UNIVERSITY OF SOUTH FLORIDA

TAMPA • ST. PETERSBURG

COLLEGE OF SOCIAL AND BEHAVIORAL SCIENCES
OFFICE OF THE DEAN
TAMPA, FLORIDA 33620

813: 974-2503

SITE: Florida Phosphate
BREAK: 17.8
OTHER: V. 54

June 10, 1975

Dr. William Hill
Polk County Health Department
Post Office Box 480
Winter Haven, Florida 33880

Dear Dr. Hill:

We are writing in reference to your participation in the "Conference on Evaluation Research in Health Care." We appreciate your interest in the Conference and your willingness to serve in the capacity we have discussed with you.

A preliminary program is enclosed. Note that all sessions are scheduled in the University Center, Room 252, on the campus of the University of South Florida. The University is located at 4202 Fowler Avenue in Tampa.

We are enclosing Xeroxed materials providing background reading. Materials prepared for the presentations will be available at the Conference.

Several motels are located on Fowler Avenue. Two of those near the University are the Holiday Inn, (813) 971-4710, and the Travelodge, (813) 971-3900. If you have any questions or difficulty relating to travel or lodging arrangements, contact us at (813) 974-2384.

Sincerely,

Travis J. Northcutt, Jr.

Lewis Bowman

TJN:LB:pje

Enclosures

EVALUATION RESEARCH IN HEALTH CARE

WITH SPECIAL APPLICATION

TO THE

FLORIDA DIVISION OF HEALTH

A Conference Sponsored

by the

College of Social and Behavioral Sciences

University of South Florida

and the

Division of Health

Florida Department of Health and Rehabilitative Services

University Center, Room 252
University of South Florida
4202 Fowler Avenue
Tampa, Florida

June 16-17, 1975

PARTICIPANTS

Dale Altman, Deputy Director
Division of Aging
1317 Winewood Boulevard
Tallahassee, Florida 32301

Lewis Bowman
Department of Political Science
University of South Florida
Tampa, Florida

Robert Browning
Division of Planning and Evaluation
Department of Health & Rehabilitative Services
1323 Winewood Boulevard
Tallahassee, Florida 32301

Michael Buscemi
Pinellas County Health Department
500 Seventh Avenue, South
St. Petersburg, Florida 33733

James Clotfelter
Department of Government
Texas Tech University
Lubbock, Texas

George Cristakis
Palm Beach County Health Department
826 Evernia Street
West Palm Beach, Florida 33402

Louis J. DiMento
Department of Political Science
University of South Florida
Tampa, Florida

John C. Doughtie
Department of Political Science
Emory University
Atlanta, Georgia 30322

Gordon Engebretson
Florida Regional Medical Program
1 Davis Boulevard
Tampa, Florida

Steven Freedman
Division of Children's Medical Services
Department of Health & Rehabilitative Services
Tallahassee, Florida 32301

Vlado Getting
Florida Technological University
Post Office Box 25000
Orlando, Florida 32816

Robert Gilbert
Department of Anthropology
University of South Florida
Tampa, Florida 33620

Lawrence Green
School of Public Health
John Hopkins University
Baltimore, Maryland

Arthur Gross
Polk County Health Department
Post Office Box 480
Winter Haven, Florida 33880

Albert Hardy
Division of Health
Department of Health & Rehabilitative Services
Jacksonville, Florida

William Hill
Polk County Health Department
Post Office Box 480
Winter Haven, Florida 33880

Godfrey Hochbaum
School of Public Health
University of North Carolina
Chapel Hill, North Carolina

James Howell
Palm Beach County Health Department
826 Everina Street
West Palm Beach, Florida 33402

Jamil Jreisat
Department of Political Science
University of South Florida
Tampa, Florida 33620

Jean Malchon
Comprehensive Health Planning Council
2400 Pinellas Point Drive, South
St. Petersburg, Florida 33712

Spero Moutsatsos
Florida Regional Medical Program
1 Davis Boulevard
Tampa, Florida

PARTICIPANTS (cont'd.)

Barbara Myers
Hillsborough County Health Planning Council
400 Courthouse Annex
Tampa, Florida 33602

Travis J. Northcutt, Jr.
College of Social & Behavioral Sciences
University of South Florida
Tampa, Florida

Patricia Paulson
Florida West Coast Health Planning Council
12945 Seminole Boulevard
Largo, Florida

Lucille Phillips
Hillsborough County Health Planning Council
400 Courthouse Annex
Tampa, Florida 33602

Eugene Piatek
Florida Regional Medical Program
1 Davis Island Boulevard
Tampa, Florida

E. Charlton Prather
Division of Health
Department of Health & Rehabilitative Services
Jacksonville, Florida

Edna Prince
Hillsborough County Health Department
Post Office Box 1731
Tampa, Florida 33601

John Prochera
Department of Political Science
University of South Florida
Tampa, Florida 33620

Herbert Traxler
Bureau of Research & Evaluation
Department of Health & Rehabilitative Service
1323 Winewood Boulevard
Tallahassee, Florida 32301

Lewis Willis
Division of Health
Department of Health & Rehabilitative Service:
Jacksonville, Florida

Curtis Wienker
Department of Anthropology
University of South Florida
Tampa, Florida 33620

Linda J. Zuro
Florida State Employment Service
800 Twiggs Street
Tampa, Florida 33602

NOTE: All sessions will meet in the University Center (Room 252) on the University of South Florida campus located at 4202 Fowler Avenue, Tampa, Florida.

June 16: Registration -- 8:30 - 9:30 a.m.

June 16: Morning Session -- 9:30 - 12:30

TOPIC: EVALUATION RESEARCH

Moderator: Lewis Bowman

Introductory remarks: Travis J. Northcutt, Jr.
Albert Hardy
Lewis Bowman

Presentation: "Conceptual, Methodological, and Practical Problems in Evaluation Research" -- Lewis Bowman, Louis J. DiMento, John S. Prochera

Coffee break: 10:30 - 10:45

Presentation: "General Design of Health Services for Evaluation" -- Godfrey Hochbaum

Commentators: Robert Browning
James Clotfelter
Spero Moutsatsos

June 16: Lunch -- 12:30 - 2:00

June 16: Afternoon Session -- 2:00 - 5:00

TOPIC: METHODOLOGIES UTILIZED IN HEALTH CARE EVALUATION RESEARCH

Moderator: Curtis Wienker

Presentation: "Utilization of Cost Effectiveness in Evaluation in the Health Care Field" -- Lawrence Green

Commenators: Louis J. DiMento
Herbert Traxler
Jamil Jreisat

Coffee break: 3:30 - 3:45

Presentation: "Evaluation as an Administrative Tool"-- Vlado Getting

Commentators: Jean Malchon
Steven Freedman

June 17: Morning Session -- 9:00 - 12:30

TOPIC: EVALUATION RESEARCH AND THE STATE DIVISION OF HEALTH

Moderator: Travis J. Northcutt, Jr.

Presentation: "Research Needs in Solving Problems in Human Services" -- Robert Browning

Presentation: "The Division of Health's Evaluation Research Needs" -- Albert Hardy

Coffee break: 10:15 - 10:30

Presentation: "Problems in Planning for the Evaluation of the Cardiovascular Risk and Status Reversal Project" -- George Cristakis

Commentators: Eugene Piatek
Curtis Wienker

June 17: Lunch -- 12:30 - 2:00

June 17: Afternoon Session -- 2:00 - 4:00

TOPIC: PRIORITIES FOR AN EVALUATION RESEARCH AGENDA FOR THE STATE DIVISION OF HEALTH

Moderator: Lewis Bowman

Perspectives for discussion:

"Local Health Officer Perspectives" (Howell, Prince)

"State Division of Health Perspectives" (Hardy, Willis)

"'Outside' Perspectives" (Traxler, Hochbaum, Paulson)

"Florida Regional Medical Program Perspectives" (Piatek, Engebretson)

"University Research and Service Perspectives" (Hochbaum, Northcutt, Prochera, Doughtie)

requested an additional funding of $350,000 for the next two years.* The requested budget made provisions for social services, proposed to study condemned persons according to their class origin, and to develop single-equation and simultaneous-equation dynamic models.

The Last Rites

A telegram from the State Capital announced that a special committee of the legislature wanted to confer with the principals of the team. The director, the operations researcher, the social worker, and the sociologist took the next available plane to the State Capital.

At the conference, the team members were happy to learn that the special committee had accepted the team's evaluation report and the request for further funding, pending clarification about the provision for social services. The special committee suggested that the budget item for social services be approved with the proviso that efforts be made to get the recipients of social service to pay for these services whenever possible.

The team came out of the conference in a jubilant mood. It was late afternoon, and they found that the next plane home was not scheduled till the following morning. They stopped in a bar for refreshment.

It was in the bar that the director of the team revealed that he had an appointment with one of the members of the special committee later that evening to discuss some business. The operations researcher decided that he would visit a relative that evening. The social worker wanted to visit a famous settlement house in the city. When the sociologist was asked what his plans were, he admitted that he was feeling rather depressed, and that he would perhaps visit some of the singles' bars. The social worker reminded him that he was, after all, a married man, that he was contemplating a violation of his "contract" with his wife, and was perhaps in need of family counseling.

REFERENCES

Blalock, H. M. *Theory construction.* Englewood Cliffs, N.J.: Prentice Hall, 1969.
Blau, P. *Exchange and power in social life.* New York: Wiley, 1964.
Campbell, D., & Stanley, J. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1963.
Goffman, E. *The presentation of self in everyday life.* New York: Anchor, 1959.
Parsons, T. *The social system.* New York: The Free Press, 1951.

* At this writing, details of the legislation passed regarding method of executing condemned persons are not available.

Evaluation Research: Possibilities and Limitations*

SHIRLEY S. ANGRIST

Evaluation research offers several benefits: scientific formulation of a policy problem, measurement of key variables, and a signal whether important social changes are occurring. Evaluation of social programs seems to answer two kinds of questions well: (a) does it matter if we choose one program over another? and (b) can we alter the number of people in a given social category? The limitations of evaluation lie in the minimal effects of most programs and in the reactive nature of both the program and the evaluation. A theory of evaluation research that indicates the possible types and amounts of impact created by various kinds of social programs, as well as the differences between long-term and short-term effects, is needed, argues the author.

Whether it is called evaluation research or policy analysis, scientific assessment of social problems has aroused the interest of two distinctive segments of the population: social scientists and their clients, the policy-makers, in all levels of government. Most notably, since 1967, federal legislation involving social programs increasingly refers to evaluation in the text of an act or in pending bills. A report prepared by the General Accounting Office (1972) reveals that much of the legislation enacted in health and safety, education, housing, and several other areas between 1967 and 1972 includes references to evaluation. A few acts specifically authorize evaluation funding and some spell out measures of effectiveness, data collection, forms of analysis, and reporting requirements.

Governmental concern with scientific assessment of national programs

* The author wishes to thank the *Journal*'s reviewers and Alfred Blumstein, Paul Brophy, William W. Cooper, Irwin Deutscher, Charles M. Eastman, and Louis Zurcher —all of whom provided helpful criticism of earlier versions of this article.

Shirley S. Angrist is associate professor of sociology, School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

is largely a culmination of several trends: 1) disillusionment with the social legislation of the early 1960s and with that bastion of American social institutions, the public schools: a widespread sense that the billions of dollars expended had yielded bottomless wells of human need with little social improvement (Piven & Cloward, 1971; Schultze *et al.*, 1972); 2) growing recognition that social scientists have bags of tricks; e.g., skills in conducting surveys, cost-benefit analysis, program planning and budgeting, computer simulation of real-world problems, and many others, with which to measure systematically the presumed intangibles of a society's operations; 3) a current shift in social scientists' interests (a frequent occurrence in the history of social science) from studying *what is* to helping design the good life—partly a response to societal changes and partly a result of federal constraints on research funding; 4) a new emphasis on social indicators that allow predicting the future, or gauging the social health of the nation, or sounding the alarm of impending riots; in other words, a push to find the social analogs of the GNP and the Consumer Price Index.

So it seems likely that henceforth social programs, small and large, in specific communities and across the nation, will be scrutinized and appraised for their effects. On the one hand, books and articles on evaluation research began appearing in the mid-1960s to codify methods, strategies, principles, and goals of evaluation. On the other hand, a few national social programs, such as Head Start and compensatory education for the disadvantaged, actually were evaluated. The cry for evaluation, accompanied by primers on how to evaluate, soon brought fresh disappointment. The rigors of tests, surveys, measurement, cost-benefit ratios, often merely dramatized the lack of effects—or what is even worse, raised accusations that the evaluative study was itself so deficient as to be worthless for policy decisions. Two examples will suffice.

Equality of Educational Opportunity, by Coleman and others (1966), evaluated the comparative benefits of public education to black and white children from the North and South. The study showed little effect of teachers, facilities, curriculum, and student body on pupil verbal achievement when family background is taken into account. But lower-class students do better when they are in school with higher performing middle class children than in all-lower-class schools (Coleman, 1972). The effects were small and the report raised clouds of methodological and substantive criticisms, including the argument that the study's limitations were so serious as to preclude policy inferences (Cain & Watts, 1972).

The poverty program led to Head Start, a large-scale program to create educational opportunities for preschool children of disadvantaged families (Westinghouse . . . , 1969). Test results on cognitive and affective measures for representative samples of Head Start children in grades 1, 2 and 3 were not appreciably higher than scores for matched control groups of non-Head Start children. The slight cognitive gains and absence of affective gains meant that Head Start had been largely ineffective in its primary goal of increasing educational opportunity for the poor (Williams & Evans, 1972).

Few evaluation studies have shown that the programs studied make an impact; in a review of 10 high quality and well-known published evaluative studies, Elinson (1972) found that none turned up much success that resulted from the program. As Weiss puts it: "Probably the most serious political problem of all is that evaluation results, with dismaying frequency, turn out to be negative. . . . To judge from evaluations, most action programs do not make much change in the behavior of individuals and groups" (Weiss, 1972c, p. 332).

For this discussion, "evaluation research" means research whose task is to assess a situation or program and to provide information for decision making. That definition includes Weiss's (1972a) notion that social science research methods are used to assess social action programs, Suchman's (1967) view that evaluation research is the study of planned change, and the view widely held by educational researchers like Alkin (1972) that evaluation means establishing decision areas, selecting, collecting, and analyzing information useful for decision-makers. It also includes the two-pronged emphasis by Rossi (1972) and Rivlin (1971) that evaluation may involve a cost-benefit approach indicating how to choose among alternative programs, or the assessment of whether specific policies achieve their intended effects, or both. The amalgam of these several definitions contains two essential ingredients: (a) the assessment or evaluation activity itself, and (b) the purpose of evaluation, which is to yield results useful as feedback for decision making about a program.

Is it true that most programs have no effects? Why cannot evaluators tap what happens? Or perhaps people and groups do not easily change? It is the purpose of this paper to review the possibilities and benefits offered by evaluation research, to indicate some limitations we must accept, and to highlight some critical research issues which remain unanswered.

WHAT EVALUATION OFFERS

At the very least, the evaluative research enterprise leads to a scientific formulation of a policy problem. Although social scientists define their work as good and important for its own sake, policymakers increasingly value and seek sound knowledge as a basis for decisions. Since we think of ourselves as an "experimenting society" (Coleman, 1971; Rivlin, 1971), there is little choice except to utilize the social science tools and knowledge already available. Despite their limitations, these resources offer partial information, and, as Coleman (1971) argues, partial information at the time needed is better than no information at all; it is clearly better than complete information which comes too late to affect policy.

The scientific approach leads to the development of measures. Since social agencies and programs frequently have fuzzy or global goals, the pressure to evaluate programs includes the need to define, operationalize, and develop measures of effectiveness; Hatry and his associates provide excellent examples of such measures (Urban Institute, 1974) for municipal services. Although useful measures may not yield good policy research, they are the *sine qua non* for conducting such research. The extensive debate stemming from the Coleman Report (Coleman *et al.*, 1966) has encouraged further research on the effects of schooling, including a search for new, more appropriate measures. Thus, Wiley and Harnischfeger (1974) indicate that quantity of schooling (in hours and days of exposure to instruction) is an important variable affecting achievement.

In addition, scientific program formulation requires a theoretical model about behavior, the economy, polity, or society. The model delimits the problem by simplifying reality enough to make evaluation manageable. When the model is applied creatively to a problem, it offers fresh insights into complex problems. Instead of accepting only social welfare or sociological models, for example, Sloan (1971) shows how psychiatric services can begin to be measured and assessed through a market model. Granted that each theoretical model carries with it assumptions and biases; still, fresh models may reveal manipulable policy variables previously ignored at the same time that they provide new interpretations of reality. Computerized simulations of a model offer a useful strategy for evaluating effects of changes in system variables; they permit pre-evaluation of a program (Blumstein, 1971).

Aside from its social scientific credentials, evaluation of specific programs signals both a need for change and the fact that change is occurring. The decision to launch a specific social program reflects recognition that "something must be done," a sense that there are problems to solve, a belief that the specified program will ameliorate those problems. The decision to conduct an evaluation of the program expresses a hope that evaluation will reveal the program's effectiveness. Commissioning an evaluation is merely tapping into an already changing situation, ripe for faster, more dramatic change. As Zurcher and Bonjean (1970) recognize, evaluation offers social scientists an opportunity for planned social intervention. In this sense, evaluation is the symptom of a larger ongoing process; it results from the momentum for change and from the urge to document and assess that change.

Evaluation research is itself a very obtrusive social process. It includes diverse activities, such as problem definition, specification of program goals, and the data collection necessary to determine what goes on. It involves people and organizations, all vested interests which are sensitive to being evaluated; they may also assist the evaluators in some aspects of the research. As in other resarch endeavors, the very system being studied is affected by the research enterprise. Zurcher (1970) describes this process as it affects the director of a poverty program who feels that his mistakes are revealed and his decisions pressured by the ongoing evaluation. In Heisenberg principle fashion, the object is altered by the very act of measurement (Newman, 1956). Thus evaluation research causes change above and beyond the social program it assesses. The evaluation perturbs the program being evaluated much like electrodes attached to the temples to obtain an electroencephalogram perturb the subject.

A housing study illustrates how evaluation research serves both as a cause and an effect of change. In the last five years government-subsidized housing has faced growing financial and social troubles. Tenant militancy, vandalism, rent delinquency, and rapidly escalating inflation contrive to threaten the very survival of public housing. These problems led the U.S. Department of Housing and Urban Development to question the effectiveness of housing subsidies and to seek alternative policies for housing the poor. Studies were commissioned on maintenance and operating costs in public housing and these confirmed the seriousness of their plight (de Leeuw, 1970; Rydell, 1970). Next HUD began sponsoring demonstration projects to test new strategies for administering public housing, such as tenant involvement, modernized management systems for local housing authorities, and variations in ownership and tenancy arrangements. The key issue then became: How well do these innovative

strategies and modernization programs work? The only meaningful way to answer such a question, of course, is evaluation research. It seems clear that the evaluation, like the demonstration program, is the direct result of a change process already in motion to revamp subsidized housing.

Now the evaluation research takes on the character of a change agent or cause once it begins. (This holds especially for evaluations which are simultaneous with the programs they evaluate; of course, *post hoc* evaluations do not have this interactive feature.) To answer the question about the effects on modernization, HUD researchers functioned as part of a team of consultant experts. The requirements of the evaluation included collecting baseline data from housing authority records, from a survey of tenants, and from interviews with tenant leaders and housing authority staff. Such contacts required the researchers to educate the concerned parties about the purposes of evaluation, to "sell" the value of the demonstration study in order to get their cooperation. Furthermore, the results of the first data collection served as feedback to tenant committees and housing authority staff indicating "how tenants see things" or "how much rent delinquency exists in this project" or "how housing authority personnel view their jobs." Thus the evaluators helped create preprogram spin-off, they raised tenant expectations for improvement, they sensitized people to record keeping for data's sake, and at various stages they informed interested parties about how things were going. In all these ways, the evaluation activities foster changes, some serendipitous and unintended.

Few social scientists would be satisfied if their craft as evaluators was merely a signal of change. Their aim is rather to assess change in order to predict outcomes and plan programs effectively. The achievements of evaluation research are usually modest, sometimes puny in meeting this aim. But we ought to recognize two types of questions that can be answered reasonably well by social research: 1) Does it matter whether Program A or Program B is implemented? and 2) How much can we increase or decrease the number of persons who use a given program, fit into a certain status, or have a given perspective?

Economic and/or Social Criteria

The economic criterion for choosing between two programs is central to a cost-benefit type of evaluation. If the two programs yield comparable benefits, then we choose the less costly of the two. The social criterion for making such a choice first involves ascertaining whether the two programs lead to different outcomes. If the key outcomes are similar, the next choice involves values or preferences: Is Program A "better" than or more "humane" or "interesting" or "enjoyable" than Program B? The ongoing educational debate about the differential effects of various teaching styles, curricula, schools, and programs reflects such social choices. Since little empirical evidence exists to show that students perform better on achievement tests or learn more under currently fashionable pedagogical conditions, such as developmental nursery schools, inquiry teaching, discussion groups, individualized instruction and open classrooms—need we conclude that nothing matters at all? Instead of centering on the "null effects" in achievement, we may conclude that several options offer the same basic learning of the three Rs but that some children (and their parents) may prefer a particular context or learning style. This argues for offering alternatives instead of abandoning them.

Results of the income maintenance experiments so far convey similar conclusions (Department of Health, Education, and Welfare, 1973). If the question is whether poor people will work less as measured by labor force participation level, dollars earned and hours worked, under a negative income tax arrangement, then the answer seems to be *no*, not substantially. Male heads of households reduced their labor supply by about 5 to 10 per cent; for the whole family, counting both spouses, the reduction was the same. Although all the economic and social ramifications of these experiments are still unknown, we might conclude that work patterns would change little from existing ones with a nationally mandated negative income tax program, and further, it might be preferred as less obtrusive and stigmatizing to the poor than current welfare subsidies.

Numbers and Attitudes

Assessing the change of the number of persons in a given category is also a manageable evaluation task. In evaluations of the effects of institutionalization on inmates or mental patients, researchers have developed useful behavioral outcome criteria: reinstitutionalization (number who return to a mental hospital, treatment setting, or to prison); recognized repetition of the deviant behavior (further manifestation of psychiatric symptoms, recidivism) is another behavioral criterion. Recent social programs which stress behavioral outcomes are seen in the provision of job training and day care services so that welfare mothers can work (Davis *et al.*, 1973; Steiner, 1971). Goals in this type of evaluation center on altering what people do and how they live rather than how they feel or

what they think. Criteria of success may include raising the percentage of gainfully employed among the poor or decreasing the numbers on local welfare rolls; or more modestly, placing a specified number of welfare mothers in jobs in each neighborhood.

Although attitudes are not perfect predictors of behavior, they provide subjective measures, indicating people's perception of their present and future condition in life, that is, their outlook on life (Deutscher, 1973; Fishbein, 1967). The survey researchers' enthrallment with attitudes comes partly from their amenability to questionnaire and interview techniques, and partly from their utility as indicators of behavior. Since attitudes encompass aspirations, preferences, and opinions, good attitudinal measures of these constructs can yield valuable descriptions about respondents.

Some evaluative research has as its primary goal the detection and measurement of changes in attitudes. We find one example in polling-type studies of voting preferences, which seek to determine whether the nation's preference for a given Presidential candidate has increased or declined over several months' time. In studies of the effect of college on students, changes in political attitudes and personality orientations have been documented as relatively long-lasting (Feldman & Newcomb, 1969). The close relationships between attitudes and life style variables—socioeconomic status and its concomitants (education, income, occupation, race, residence), health, age, and marital status (Wilson, 1967)—all corroborate the utility of attitudes as measures of people's condition. A social program might be considered effective if it lowers negative views of one's life circumstances, while it elevates positive ones. Thus, the program goal might be to raise hopes for the future or educational aspirations for one's children, or to decrease meaninglessness or to lessen the fear of crime.

BUILT-IN CONSTRAINTS AND HAZARDS

Despite all our social science tools for assessing effects, we may have to live with the constraint that most programs have little impact. Perhaps human nature is so flexible and programs or "treatments" so similar that their outcomes are similar. Or alternatively, the time frames both for the programs and their evaluations may be too short. The urge for "quick" answers, the pressure to make informed decisions, the danger that program monies and interests will wane, all conspire to curtail program life.

Furthermore, long-term assessment, especially experimentation, is costly and difficult to set in motion. And program architects must weigh the utility of possibly weak long-run answers against short-run answers that are only somewhat weaker. Would Head Start have improved school achievement greatly if the program lasted longer? Would labor force participation decline dramatically if income maintenance became a national policy? Unfortunately, these questions lack answers.

Politicization and Natural Disaster

More dilemmas arise when longer-lasting programs are evaluated. One inevitable hazard is the "politicization" of the program and its evaluation (Weiss, 1972c). There is more time and opportunity for vested interest groups and persons to mobilize for or against the program—to help its success or seek its demise.

A second inevitable contingency is historical. Events, both routine and dramatic, may minimize, obscure, or upset the program's own natural history. Some exogenous occurrences threaten an ongoing program while others may accentuate and nourish the program. The New Jersey income maintenance experiment was affected in 1969 by the state's offer of Aid to Families with Dependent Children whose father was unemployed or underemployed. Certainly, experimental designs reduce the likelihood of historical and political impingement on the program. But they are no guarantee against differential disturbances to the experimental and control groups. For example, in matched pairs of housing projects one may be torn down or ravaged by fire, with no control group left. Nor can experimental designs avoid the consequences of political maneuvers or subversions which devastate the program.

Titrating the pros and cons of long and short programs, no easy answer emerges. But settling for six-month or one-year experiments in education, health, housing, and the like suggests naivete about social processes and individual behavior. Preparing for events and organizing people around tasks take time. But social scientists have seldom studied the precise relationship between duration of events and their impact on participants.

Competing Variables and Side Effects

We may also have to live with the inevitable search for the pertinent variables. One explanation for finding "no effects" is that the evaluators assessed the wrong variables. The educational sphere has witnessed a tirade against achievement tests as inappropriate or narrow measures of

what educational programs accomplish (Shapiro, 1973; White & Larsen, 1972). Evaluators then build a "chain of heroic assumptions" on such measures (Ribich, 1968; Rivlin, 1971). This interpretation of the problem would be plausible if there were evidence that other effects do occur even though test performance shifts little. The evidence is strong on suggestion and weak on empirical data. But it is worth considering because of its *prima facie* validity. Typically, program directors, participants, and even evaluators perceive that "things are happening"—for example, that people are happier, more involved, learn a lot, or that many formerly unreachable people have been served by the program. Shapiro (1973) raises this issue about Follow Through classrooms and their control groups: She reports the contrast between systematically observed "dramatic differences in classroom behavior" and "null differences in test behavior." If this is so, it is a case of neglect rather than ignorance, because social scientists are usually sensitive to the nature of their dependent variables as attitudes or behavior, as aggregated or disaggregated variables, as individual or group measures, and as multivariate.

Along with the "wrong variable" problem comes the related issue of side effects and serendipitous events. Centering on a simple outcome variable obscures the multiple levels of occurring events and the potential for several kinds of impact (Deutscher, 1973). It may be important to ascertain whether a program is functioning and to monitor its course without detailed concern for what it accomplishes. Monitoring program development is the simplest level of evaluation. It refers to actual implementation of the specified program and its elements. Much demonstration research remains at this outcome level. Once plans are made, their implementation becomes the key concern. The program may be judged effective by merely being carried out. It may also be considered effective if it generates spontaneous or spin-off programs. But the program goals may be to create activities, organizations, buildings. The evaluative procedure is close to but not exactly what Suchman (1967) calls "effort." It deals primarily with monitoring the process of carrying out the plans, recording the type of activity and its historical evolution. The criterion of success is not "does the program work?" but "does the program exist and function?"

Monitoring a program usually reveals a "social chain of effects" (Hyman & Wright, 1967). There are possibilities for contamination by program staff interacting with participants and difficulty in identifying the boundaries of the program being evaluated. The problem is poignantly

expressed by Goldstein, Hull, and Ostrander (1973), who show that evaluators can be engulfed by people, program, and situations, sometimes exploited, often accomplices in events beyond their control. Yet this confusion of evaluation effects with program effects need not be negative. Program participants may well benefit from the wave-like spreading and proliferation of effects. This has led Scriven (1972b) to propose "goal-free evaluation." Although this view is not widely accepted, it may be useful to conceive of the inevitable reactive effects as legitimate parts of the program's accomplishment. Even when decision-makers fail to specify goals, the evaluators must "unearth and study consequences that have significant impact on people and systems" (Weiss, 1972b).

EMERGENT EVALUATION THEORY

The beginnings of evaluative theory are in sight. We have but to consider the several recent outstanding attempts to collect and synthesize thought and practice in evaluation research—by Wholey and others (1970), by Weiss (1972a), and by Rossi and Williams (1972). They move us away from the either-or polarities of sticking with descriptive and unique case studies (Weiss & Rein, 1972), on the one hand, or conducting only true experiments (Suchman, 1967), on the other. One principle proposed by Rossi (1972) is: if no effects emerge when a "soft" evaluation method is used, no effects are likely with "harder" methods. Thus one can eliminate ineffective programs through use of "soft" methods. In a study of utilization of government-funded evaluation research, Rieker (1972) formulates hypotheses to indicate the conditions under which evaluation research is likely to be widely disseminated and utilized for policymaking. It will require tested generalizations before we can speak of evaluation theory. But we already have a few working principles along with countless cases to serve as the data for theory building.

In the meantime, we continue to strive for the best in evaluation (controlled experimentation) and accommodate to the worst (politicization). But we need research on some key matters: 1) what kinds of effects can various programs be expected to have? and 2) what differences are there between long-term and short-term programs and their effects?

In evaluative research, as in social science research generally, there are often "no effects." Yet there is an intriguing and explicit sense that changes do occur in individuals, groups, situations, and in the larger society. It is clear that the campus and urban unrest of the 1960s has

dwindled. These changes are easily measured in terms of a count of riots, sit-ins, and confrontations per month. Or consider the Uniform Crime Reports and note that the rise in crime rates declined for the first time in 1973 compared with previous years and then rose again in 1974. Those are detectable changes. Our dilemma in evaluation research is not the lack of effects so much as the rarity of intended, expected, or measurable effects. We may then despair and conclude that nothing helps, no remedy cures, no program changes people, no spending is justified. Yet as social scientists we well know that specific social contexts do have noticeable consequences. We take for granted (at least in teaching undergraduates) that social classes, ethnic, and religious groups differentially shape their members; that prisons and mental hospitals mold their inmates; that boarding schools, colleges, and sororities influence their recruits; that organizational structures constrain their functionaries. The issue is to specify what those changes are, how much change occurs, and whether the changes were anticipated.

Evaluation Continuum or Feedback Loop

Among the tasks for a theory of evaluation research, then, is clarification of the amounts and levels of impact that programs can have. The simplest impact involves the mere existence of the program. Or program planners may seek attitudinal changes in designated population segments without any behavioral requirements; for example, that tenants be more satisfied with the housing maintenance but not that they lower complaints about maintenance. Another level may involve behavioral change in whole categories of individuals. And a still further level would be system change providing evidence that the combination of certain inputs and processes yields predictable outputs.

The four levels of impact range from monitoring, as the simplest form of evaluation, to system change, the most complex. It seems likely that they form a continuum in terms of cost as well as in difficulty of execution. Since most social programs are complex, with several aims, they probably involve all four levels of impact. But evaluations which can assess behavioral and system change seem the most likely to serve as inputs to social policy.

The task of conducting sophisticated and rigorous policy research remains. Even as we witness the conduct of impressive social experiments involving housing allowance, educational vouchers, and income maintenance, the issue of effects hovers: How much impact is enough for a

policy decision? Can we reach a magnitude of difference with a confidence level strong enough to recommend a policy decision? We may face the irony that well-defined programs with precise outcome measures and experimental designs have less effect than fuzzy programs and vague measures and nonexperimental designs (Blumstein & Cohen, 1974). We also face the newness of scientifically prescribed large-scale social experimentation. Rivlin (1974) suggests that programs to change monetary incentives geared to individuals, families, or firms yield more measurable outcomes (e.g., dollars earned, hours worked, doctor visits) than services rendered to people with complicated treatments and outcomes (e.g., changes in health status or educational competence).

Time Frames and Spin-Offs

Even if we were satisfied that program effects are real, we still need to understand the time dimension of social programs. How long does it take for effects to occur? Unfortunately, we have little precise knowledge about how soon social entities carry out their mission and exist in a more or less stable form. The difficulty is that "social experiments may simply take too much time. Many of the really interesting effects of social action show up only after a period of years" (Rivlin, 1971, p. 117). Despite this difficulty, we have probably erred too far with programs of short duration and with delimited time frames.

One strategy is to opt for longer programs whenever that makes sense. Although an inoculation program may be successful within six months, testing a new health care delivery system may require two or three years. But the matter of time frame is more than an argument in favor of long-lasting programs. As described earlier, the program is itself an outgrowth of a climate for change or experimentation in a given arena. And as the program proceeds it may create spin-offs and proliferation of effects.

Several conclusions follow from these time-related features of social programs. It seems reasonable to use a wider time band to demarcate the program's beginning and end. The "before" condition should come earlier than such bureaucratic niceties as "the day the funding began" or "the end of the fiscal year." The "before" time period needs to precede totally the set of activities which construct, plan, design, and publicize the program-to-be. Similarly, the "after" condition needs defining so that the program and all its offshoots are completely over. A degree of uncertainty about locating the program's termination suggests that some period

of grace or cooling off after the official end is advisable—perhaps 6 months or a year.

Since the spin-offs are real and in a sense result from the program, the best tactic is to recognize them as early as possible, define them as effects, and ensure their accountability in the evaluation. To separate program effects from proliferation effects seems unachievable—but worse, risks throwing out something that is part of the program accomplishments. The benefits that turn up may be unanticipated; those expected could fail to appear. Evaluators need an openness to such later developments and to new, more appropriate criteria, which may emerge during the course of the program (cf. March, 1972).

Thus, to the argument for expanding the time frame, we must now add extension of the effects. Instead of holding tightly to an easily operationalized, narrowly defined single outcome measure (like IQ test score), with all its limitations reviewed earlier, we need to consider systematically the range of likely effects and find ways to gauge them.

If we recall that the purpose of evaluative research is to yield outcomes which lead to action, e.g., given outcome x, then policy y should be followed; then formulation of decision rules must accompany the process of establishing program goals. When such rules are not part of the program planning process, evaluative research can serve only limited usefulness. It can yield a summative evaluation (Scriven, 1972a): was the goal achieved by the program? This yields a product view of the program outcome. A process view requires that evaluation be formative, embedded in the program planning; the more immediate outcomes then serve to make corrections, alterations, or total revisions in the program. Ideally, then, evaluation is part of an ongoing feedback loop, in which program outcomes are judged against program goals; if the two do not fit, the program is modified to result in different outcomes in order to achieve the program goals.

The potentially wide scope of program outcomes can incorporate both formative and summative evaluations. This view seems to have taken hold partly because of the difficulties in conducting tidy evaluations but partly in recognition that the input–output model is not enough, that we need to keep track of what happens in the black box. Instead of the educational researchers' term "formative evaluation," behavioral scientists refer to "social audit" (Coleman, 1971), "social reporting" (Colantoni, Cooper, & Dietzer, 1972), a "natural history" account (Caro, 1969), or a "concurrent evaluation" (Lazarsfeld, Sewell, & Wilensky, 1967). The idea

common to these disparate terms is description of the process with feedback useful to decision-makers for program correction. This view makes possible the combination of long-term impact or summative evaluation with short-term, ongoing adaptive changes.

This paper offers small remedies for large problems. Furthermore, they are not mutually exclusive. Unfortunately there is no apparent solution to the complexities which the evaluators must evaluate. Although the state of the art of evaluation research remains more primitive than elegant laboratory experiments and definitive cost-benefit ratios, we are clearly past the case study that is unique in people, program, time, and place. We are ready for generalized theory with clear applications.

The big remedy, then, would be theory that synthesizes the accumulated information, data, experience, and methodologies to allow early specification of the conditions under which given programs will work or will founder; which kinds of measures are appropriate; what consequences stem from the inevitable politicization of the program; how individuals and aggregates are likely to react to various strategies; how long it takes for programs to reach some equilibrium stage or to become established; the conditions which favor long-term effects; the trade-off in costs and benefits from alternative evaluation strategies; the comparative value of outside evaluators and in-house evaluators; patterns of working relationships among funding agencies, program planners, directors, participants, evaluators, and the public; when to hold out for a classical experiment and when to seek other research designs. The growing interest in program assessment and the accumulating body of evaluation research should facilitate the emergence of such theory.

REFERENCES

Alkin, M. C. Evaluation theory development. In C. H. Weiss (Ed.), *Evaluating action programs*. Boston: Allyn and Bacon, 1972. Pp. 105-117.

Blumstein, A. Systems analysis and planning for the criminal justice system. Unpublished paper, School of Urban and Public Affairs, Carnegie-Mellon University, 1971.

Blumstein, A., & Cohen, J. An evaluation of a college level program in a maximum security prison. Unpublished report, School of Urban and Public Affairs, Carnegie-Mellon University, 1974.

Cain, G. G., & Watts, H. W. Problems in making policy inferences from the Coleman Report. In P. H. Rossi and W. Williams (Eds.), *Evaluating social programs*. New York: Seminar Press, 1972. Pp. 73-95.

Caro, F. G. Approaches to evaluative research: A review. *Human Organization*, 1969, 28, 87-99.

Colantoni, C. S., Cooper, W. W., & Dietzer, H. J. Budgetary disclosure and corporate social accounting. Paper presented at the Conference on Corporate Social Accounts, Seattle, 1972.

Coleman, J. S. Problems of conceptualization and measurement in studying policy impacts. Paper presented at the Conference on the Impacts of Public Policies, St. Thomas, U.S. Virgin Islands, 1971.

Coleman, J. S. Reply to Cain and Watts. In P. H. Rossi and W. Williams (Eds.), *Evaluating social programs*. New York: Seminar Press, 1972. Pp. 97-107.

Coleman, J. S., Campbell, E. O., Hobson, C. E., McPartland, J., & Mood, A. M. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966.

Davis, O. A., Doyle, P. M., Joseph, M. L., Niles, J. S., & Perry, W. D. An empirical study of the NAB-JOBS program. *Public Policy*, 1973, 21, 235-262.

de Leeuw, F. *Operating costs in public housing*. Washington, D.C.: The Urban Institute, 1970.

Department of Health, Education, and Welfare. *Summary report: New Jersey graduated work-incentive experiment*. Washington, D.C.: H.E.W., 1973.

Deutscher, I. *What we say/what we do*. Glenview, Ill.: Scott, Foresman, 1973.

Elinson, J. Effectiveness of social action programs in health and welfare. In C. H. Weiss (Ed.), *Evaluating action programs*. Boston: Allyn and Bacon, 1972. Pp. 294-299.

Feldman, K. A., & Newcomb, T. M. *The impact of college on students*. San Francisco: Jossey-Bass, 1969.

Fishbein, M. *Readings in attitude theory and measurement*. New York: Wiley, 1967.

General Accounting Office. *Legislative references to evaluation 1967-1972*. Washington, D.C.: G.A.O., 1972.

Goldstein, P., Hull, D., & Ostrander, S. The evaluation of urban ameliorative projects: Case studies and integrating framework. Paper presented at the meeting of the North Central Sociological Association, Cincinnati, 1973.

Hyman, H. H., & Wright, C. R. Evaluating social action programs. In P. F. Lazarsfeld, W. Sewell, and H. Wilensky (Eds.), *The uses of sociology*. New York: Basic Books, 1967. Pp. 741-784.

Lazarsfeld, P. F., Sewell, W., & Wilensky, H. (Eds.) *The uses of sociology*. New York: Basic Books, 1967.

March, J. G. Model bias in social action. *Review of Educational Research*, 1972, 42, 413-429.

Newman, J. R. Commentary on uncertainty. In J. R. Newman (Ed.), *The world of mathematics*, Vol. II. New York: Simon and Schuster, 1956. Pp. 1047-1051.

Piven, F. F., & Cloward, R. A. *Regulating the poor: The functions of public welfare*. New York: Vintage Books, 1971.

Ribich, T. I. *Education and poverty*. Washington, D.C.: The Brookings Institution, 1968.

Rieker, P. The utilization of evaluation studies. Paper presented at the meeting of the American Sociological Association, New Orleans, 1972.

Rivlin, A. M. *Systematic thinking for social action*. Washington, D.C.: The Brookings Institution, 1971.

Rivlin, A. M. Policy analysis and large scale social experiments. Paper presented at the 14th Carnegie Conference on Large Scale Social Experiments, Carnegie-Mellon University, 1974.

Rossi, P. H. Testing for success and failure in social action. In P. H. Rossi and W. Williams (Eds.), *Evaluating social programs*. New York: Seminar Press, 1972. Pp. 11-49.

Rossi, P. H., & Williams, W. (Eds.) *Evaluating social programs: Theory, practice and politics*. New York: Seminar Press, 1972.

Rydell, C. P. *Factors affecting maintenance and operating costs in federal public housing projects*. New York: Rand Institute, 1970.

Schultze, C. L., Fried, E. R., Rivlin, A. M., & Teeters, N. H. *Setting national priorities: The 1973 budget*. Washington, D.C.: The Brookings Institution, 1972.

Scriven, M. The methodology of evaluation. In C. H. Weiss (Ed.), *Evaluating action programs*. Boston: Allyn and Bacon, 1972. Pp. 123-136. (a)

Scriven, M. Pros and cons about goal-free evaluation. *Journal of Educational Evaluation*, 1972, 3, 1-8. (b)

Shapiro, E. Examining criteria for evaluating educational programs. Paper presented at the meeting of the Society for Research in Child Development, Philadelphia, 1973.

Sloan, F. A. *Planning public expenditures on mental health service delivery*. New York: Rand Institute, 1971.

Steiner, G. Y. *The state of welfare*. Washington, D.C.: The Brookings Institution, 1971.

Suchman, E. A. *Evaluation research*. New York: Russell Sage, 1967.

Urban Institute. *Measuring the effectiveness of basic municipal services*. Washington, D.C.: the Institute, 1974.

Weiss, C. H. (Ed.) *Evaluating action programs: Readings in social action and education*. Boston: Allyn and Bacon, 1972. (a)

Weiss, C. H. *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, N.J.: Prentice-Hall, 1972. (b)

Weiss, C. H. The politicization of evaluation research. In C. H. Weiss (Ed.), *Evaluating action programs*. Boston: Allyn and Bacon, 1972. Pp. 327-338. (c)

Weiss, R. S., & Rein, M. The evaluation of broad-aim programs: Difficulties in experimental design and an alternative. In C. H. Weiss (Ed.), *Evaluating action programs*. Boston: Allyn and Bacon, 1972. Pp. 236-249.

Westinghouse Learning Corporation—Ohio University. *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development*, 1969.

White, S. H., & Larsen, R. E. The evaluation of early childhood education. Paper presented at the meeting of the American Sociological Association, New Orleans, 1972.

Wholey, J. S., Scanlon, J. W., Duffy, H. G., Fukumoto, J. S., & Vogt, L. M. *Federal evaluation policy*. Washington, D.C.: the Urban Institute, 1970.

Wiley, D. E., & Harnischfeger, A. Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher*, 1974, 3 (4), 7-12.

Williams, W., & Evans, J. W. The politics of evaluation: The case of Head Start. In P. H. Rossi and W. Williams (Eds.), *Evaluating social programs*. New York: Seminar Press, 1972. Pp. 249-264.

Wilson, W. Correlates of avowed happiness. *Psychological Bulletin*, 1967, 64, 294-306.

Zurcher, L. A., Jr. *Poverty warriors: The human experience of planned social intervention*. Austin: University of Texas Press, 1970.

Zurcher, L. A., Jr., & Bonjean, C. M. *Planned social intervention: An interdisciplinary anthology*. Scranton: Chandler, 1970.

Soc. Sci. & Med., Vol. 9, pp. 133–139. Pergamon Press 1975. Printed in Great Britain.

BENEFIT-COST ANALYSIS: A REVIEW OF ITS APPLICABILITY IN POLICY ANALYSIS FOR DELIVERING HEALTH SERVICES*

DAVID W. DUNLOP

Health Care Administration and Planning Division, Meharry Medical College; and Department of Economics, Vanderbilt University, Nashville, Tennessee, U.S.A.

Abstract—The use of benefit-cost analysis for programmatic decision making in health is reviewed focusing on the kinds of health programs analyzed, the impact on health policy, and the problems encountered when using such an analytical tool. Alternatives to benefit-cost analysis are discussed in light of their policy relevance. A concluding section provides a broader perspective for health planning and program development.

Over the last decade increasing use has been made of the analytical methodology known as benefit-cost analysis for public policy determination where limited governmental resources have necessitated the use of systematic methods for rationalizing the efficacy of new or continued public support of a number of social and economic programs. Health and medical care programs have not been overlooked by policy-makers in their attempts to utilize benefit-cost analysis for making programmatic and resource allocation decisions. In the United States, under the prodding of President Johnson, the methods of planning, programming and budgeting, which incorporated the use of benefit-cost analysis, became widely used in the late 1960's in health program analysis and continue to be used in a variety of ways today [1]. Increasingly, other countries throughout the world have initiated efforts to improve the delivery of health and medical care by planning the expansion of health delivery systems and evaluating programs which are designed to eradicate disease and hunger and improve the quality of life [2]. Benefit-cost analysis has been incorporated into such planning activities as one of the corner stones upon which resource allocation decisions have been based. In this paper, the use of benefit-cost analysis for programmatic decision-making in health throughout the world is reviewed. The review focuses particularly on: (a) the kinds of health programs which have been analyzed; (b) the impact of such analyses on policy decisions; and (c) the problems encountered by those attempting to use benefit-cost analysis. The second section of the paper focuses on the "benefits and costs" of alternative methodologies to benefit-cost analysis for health policy decision-making. A concluding section provides a broader perspective for health planning and programmatic development.

REVIEW OF THE USE OF BENEFIT-COST ANALYSIS IN HEALTH

The kinds of health programs analyzed

Benefit-cost analysis has generally been used to analyze health investments in one of three ways. First, and perhaps most common, it has been used to analyze disease-specific programs of intervention. Among the most common diseases analyzed in this way are malaria [3], heart disease [4], polio [5], syphilis [6], and kidney disease [7]. Where specific diseases have been so analyzed, the analysis has typically focused on the economic impact of the disease, measured in terms of the cost of premature death and disability due to a reduction in the capacity to function, the general assumption being that the opportunity cost of such disease-specific events is the age and sex-specific average annual earnings of those so implicated. While most studies have referred to other social and economic benefits, particularly that of the cost of debility [8] from a specific disease, no systematic attempt has been made to operationalize this concept and estimate its relative magnitude. Finally, it is important to mention that virtually all of the analyses conducted on disease-specific intervention programs have been prospective, or *ex ante* in nature. While such analyses may be useful to the public policy-maker who must decide whether to initiate a certain health investment, seldom are such analyses conducted on an *ex post* basis for public reassessment of initial decisions.

Benefit-cost analysis has also been used in the health field to analyze a set of alternative ways in which a given health or medical care service can be produced or distributed. (In many situations the resulting analysis is similar to cost-effectiveness analysis which is discussed subsequently in the paper.) One of the more interesting studies of this type was conducted by the U.S. Department of Health, Education and Welfare on alternative ways of reducing motor vehicle deaths and injuries [9]. In this particular study, a number of potential interventions were analyzed to determine which programmatic mix had the highest benefit-cost ratio.

Thirdly, benefit-cost analysis has been used to estimate the return to public investment in health-related programs such as health and medical research [10] and health manpower training programs [11]. While not belittling such analyses, it is interesting to observe that virtually all studies conclude that the benefits to health investments outweigh the cost, whether couched in terms of a rate of return, benefit-cost ratio, or net present value.

* Delivered to the 4th International Conference on Social Science and Medicine, Elsinore, Denmark, 1974.

IMPACT OF BENEFIT-COST STUDIES IN HEALTH ON POLICY DECISIONS

While it is presumptuous to assume that a benefit-cost study provides a public policy-maker with sufficient justification to make a particular decision, there is evidence to suggest that such studies have played an important role in public policy determination. Perhaps the most widely known incident in which a benefit-cost study was implicated in policy determination, was in the area of family planning and birth control—a programmatic area with substantial health overtones. It was not long after Stephen Enke had reported his findings that the benefits of birth control were approximately 100 times the cost that President Johnson commented on these findings and the U.S. Congress significantly expanded U.S. AID funds to assist developing countries in implementing such programs [12].

In the development and expansion of flouridated water supplies in the U.S., a benefit-cost study has nearly always preceeded the development of a dental health program in a previously unflouridated area of the country. The results are usually widely disseminated to the public, just prior to the election where a public decision is made [13].

In addition to these two examples where public policy in health has been influenced by the findings of benefit-cost analysis, there are at least two other comments which should be made. First, much of the benefit-cost analysis conducted on a disease-specific basis, whether in the United States or elsewhere, has been done with an eye toward what might be called the "preventive vs curative" debate. Most studies conclude that it is economically inefficient to continue to expand curative systems of health care for specific diseases where there is a preventive technology available.

Second, most public policy recommendations emanating from benefit-cost studies have been of two types: (a) start a new health program to address a particular health problem; or (b) expand an existing one in order to achieve the potential benefits identified with intervention strategy. It is disturbing that little use has been made of analytical techniques, such as benefit-cost analysis and other evaluation methods, to address the qualitative rather than purely quantitative aspects of public policy as it relates to health [14]. In the concluding section of this paper, an attempt is made to provide constructive suggestions related to this point.

PROBLEMS ENCOUNTERED IN USING BENEFIT-COST ANALYSIS IN EVALUATING HEALTH PROJECTS

Several problems have been particularly difficult to overcome in attempts to apply benefit-cost analysis to a particular health or medical care investment. These problems may be classified according to the following taxonomy: problems of (a) the theoretical underpinnings of benefit-cost analysis; (b) unresolved issues of the methodology of benefit-cost analysis; (c) the implicit objective function or decision criteria which are incorporated in the outcome indicators, i.e. rate of return, net present value, and the benefit-cost ratio, and (d) issues of measurement and outcome conceptualization specific to health and medical care investments.

The first problem—the theoretical underpinning of benefit-cost analysis—is related to the fact that the analytical framework of benefit-cost analysis is a partial framework and is basically a micro-oriented tool. As long as it is used to analyze economic problems with marginal impacts upon the entire economic, social and demographic structure of a society, it may be appropriately used. Such would be the case where the analysis is applied to alternative production technologies for the delivery of a particular health service or in determining the social "payoff" to a concerted effort designed to reduce the incidence or eradicate a specific disease which is not a major cause of mortality, morbidity or debility in the population. More succinctly stated, benefit-cost analysis assumes no changes in prices or quantities of the supply of factors of production or in the pattern of demand for a given set of health services over time as a consequence of a particular health or medical care investment. Unfortunately, this basic micro-structural assumption underlying the benefit-cost analytical framework makes it particularly vulnerable when applied to health problems with a potential for a multiplicity of interactions with other societal institutions, not the least of which being its demographic structure. The problem of extra-marginal impacts is particularly acute in developing countries where the scourge of infectious and parasitic diseases is so great. If malaria were to be eradicated, for example, there would be major demographic impacts felt throughout the society, not necessarily in the short run, but during the subsequent 20–30 yr. Without a more sophisticated analytical framework capable of accounting for such macroeconomic impacts, the public policy decisions based on the results of a micro-economic benefit-cost framework can be undesirable [15]. A similar situation is possible in more developed countries if the only analysis made of the social and economic impacts of a significant reduction in the incidence of heart diseases, for instance, is conducted using a benefit-cost framework. Similarly, if there were interest in analyzing the economic impact of a significant technological breakthrough in the organization and/or delivery of health and medical care in the United States, a benefit-cost framework would likely be inappropriate due to the extra-marginal impacts such a change would have on the structure of the American economy.

The second problem—unresolved issues of methodology in using benefit-cost analysis—is endemic to all studies using the approach, and is not necessarily unique to health. Although these problems have been adequately covered in the literature [16], it is important to mention them briefly here. The first problem is a conceptual one of determining what to count as a benefit and what to count as a cost. Conventions are developing on this issue and for the most part can easily be followed. A second and more difficult problem is the issue of the appropriate discount rate to use in comparing the costs and benefits over time. The question becomes one of determining society's time-rate of preference [17]. As a practical matter, several alternative rates of interest have been employed in conducting sensitivity analysis of the results to determine the extent to which project ranking may change as a result of different rates of interest being employed. Third is the problem of determining the appropriate measure to use in ranking projects. Should one calculate and

rank projects on the basis of: (a) internal rate of return, (b) benefit-cost ratio, or (c) net present value (NPV)? It is generally agreed that the most appropriate ranking criterion, assuming that the objective function underlying the investment is to maximize the return to society (social profit, if you will), is the NPV measure [18]. The problem with using this measure, rather than one of the other two, however, is that it is not as easy for public policy-makers to intuitively interpret its significance.

The third basic problem with benefit-cost analysis is that it implicitly assumes that the objective underlying a decision to invest in health or medical care programs is to maximize economic growth by maximizing the return to society's investment. All three outcome indicators or measures of benefit-cost analysis (rate of return, benefit-cost ratio, or NPV) provide the public policy-maker with information as to the relative efficacy of a particular health investment in obtaining the target of maximizing economic growth, assuming the present price-structure adequately reveals society's preferences. However, if public policy-makers have additional objectives they would like to consider in making investment decisions, such as equity and distributional considerations, then it becomes necessary to incorporate them into the analysis. At this juncture, such methodological improvements in the framework of benefit-cost analysis are beginning to be made, with the general approach being to disaggregate the benefit-cost analysis of a proposed investment and conduct it on a group- or geographic-specific basis [19]. As long as considerations of equity and distribution are to be given weight by policy-makers in health and medical care investment decisions, improvements in benefit-cost analysis must be forthcoming and reliance on single indicators such as NPV, etc., to rank projects for decision-making purposes must be reduced.

The last problem which has been particularly difficult for economists and others using benefit-cost analysis to deal with in evaluating health programs has been in (a) conceptualizing an operational output measure of a health program and (b) obtaining the relevant statistical support to estimate the costs and benefits of a particular health investment. While some attempt has been made to overcome the conceptual difficulties of describing appropriate outcome measures in both preventive and curative health and medical care programs [20], the development of appropriate statistical indicators amassed on a regular basis is still generally lacking [21]. Perhaps one of the most underrated benefits of the increased public pressure to intervene in the delivery of health services in the U.S. will be the greater attention paid to amassing more complete and appropriate statistical information related to the costs and benefits of specific health investments [22].

ALTERNATIVES TO BENEFIT-COST ANALYSIS IN ANALYZING HEALTH AND MEDICAL CARE INVESTMENTS

There are three alternative methodologies which may be usefully employed in analyzing the economic and social effects of health and related investments, including investments designed to restructure the delivery system of health services. These alternatives include: (a) cost-effectiveness analysis (C-E); (b) linear programming and operations research methods; and (c) systems analysis, including simulation. A brief review of the advantages and disadvantages of these three alternatives follows below.

Cost-effectiveness analysis

Cost-effectiveness analysis is similar to benefit-cost analysis in that it is a micro-oriented tool of analysis. However, it has more flexibility than benefit-cost analysis as an analytical tool, since it is not circumscribed by reducing all effects into value terms by using a common denominator such as money [23]. This presents the analyst and policymaker with the opportunity to consider a set of possible outcomes from a given investment or programmatic change. Assuming that there is an adequate measure of the resources used in the project and that they have been valued in such a way as to reflect their respective opportunity costs, one can estimate the cost per unit of effect and disaggregate the analysis on an effect-specific basis. The virtue of this approach for health program decision-making is that qualitative indicators can be considered in conjunction with quantitative measures of effects. Once one has estimated the cost per unit of effect, one can estimate—by letting one of the effects become a numeraire—a set of price ratios between effects to enable policy-makers to estimate the trade-off between various effects.

This methodology is particularly useful where the objective of the analysis is to determine the most appropriate alternative strategy for obtaining a set of agreed upon effects. In such a situation, the analysis reduces to a search for the cost minimizing strategy. In health, this type of problem may easily surface, for instance, in situations where one is evaluating alternative immunization technologies for a measles campaign to protect at least 90 per cent of the children in the population within a two-week period. In addition, it may be effectively used to evaluate alternative health manpower training programs which have a certain target *cadre* of persons to train.

Finally, the methodology of cost-effectiveness can be beneficially incorporated into the planning, programming and budgeting process which is increasingly being used by all government agencies as a systematic way of continuously reviewing programmatic activities. It can, for example, be utilized as the principal evaluation technique in determining which set of administrative procedures within the Ministry of Health was most effective in eliciting new ideas for improving the delivery of a particular type of health service administered and financed by the Ministry. In addition, since the analytical process implied by cost-effectiveness analysis requires it, additional information is obtained in a routine manner for purposes of public policy and administrative review.

Cost-effectiveness analysis has at least two disadvantages, however. First, it is difficult to apply discounting techniques to the flow of effects of a health program over time. Let us assume, for example that one were using a cost-effectiveness framework to analyze three alternative technologies for immunizing a specified population. If the primary indicators of effectiveness were a reduction in disease-specific mortality and morbidity, how would one evaluate two alternatives: one in which there was a relatively minor reduction in both indicators during the first time period, and one in

which the effects were major reductions in both morbidity and mortality, but where such effects would not occur for 5-10 yr? Without employing discounting techniques and a common valuing scheme, it would be difficult to make meaningful comparisons between programs. Such difficulties would be heightened where the set of effects of each alternative being analyzed included qualitative indicators as well as quantitative ones. For example, changes in affective rather than cognitive learning with respect to substance abuse (drugs, alcohol, tobacco, etc.), which have health status impacts, may be the objective effects of a particular preventive mental health program. At the present time, societies throughout the world have difficulty rationally discussing such issues, let alone measuring, with any degree of accuracy, effects of such programs, taking into consideration the time path of those effects.

The second problem with cost-effectiveness as an analytical tool for evaluating health and medical care programs arises when it is necessary to analyze two or more programs, projects or production alternatives having differing sets of possible effects. For example, the methodology cannot deal with comparing the effects of a health program which has the objective of reducing the rate of infant mortality with a program designed to reduce the incidence of deaths and injuries from auto-accidents. This problem dramatizes the fact that cost-effectiveness analysis, like benefit-cost analysis, is a micro-oriented tool and is appropriately employed in analyzing situations where the effects of alternatives are similar.

Linear programming and operations research

Linear programming techniques have been increasingly used to analyze certain problems in health delivery systems. Perhaps they have been used most effectively in analyzing the effects of production decisions made at the health firm or institution level (hospital, doctors' office, health center, etc.) on the set of health services produced by that unit [24]. A further normative evaluation is then required to determine whether such implied output changes are desirable from a private or societal point of view. Where an entire health services delivery system is homogeneous in terms of the resource endowment, the production techniques employed (such as in the diagnostic and treatment phases of service delivery) and in the technology used to produce each service, linear programming may also be used to analyze the impact of policy changes at a more aggregative level [25].

There are several problems with such techniques, however. First, production relationships must be assumed to have a linear form such that there are fixed coefficients between inputs and outputs. This problem is generic to the Linear Programming Technique [26]. Second, linear programming technique requires that the desired outcome be specified in unambiguous terms, using a single outcome measure, i.e. profits or output. This problem creates considerable difficulty for health decision-makers and social planners since multiple objectives, such as efficiency and equity, may be simultaneously under consideration. As in benefit-cost analysis, linear programming models can be used on a disaggregated basis to allow for equity considerations. However, this is a second best approach to the problem.

A third problem with using linear programming techniques in the health field has been a conceptual one of output specification. Succinctly stated, the question is: what does a health delivery system produce? Efforts have been made to answer that question in an operationally useful way [27], but it remains a continual problem which must be addressed in each specific application of linear programming to the health field.

Finally, in order to use linear programming and operations research techniques, it becomes essential to have a well-developed data base, with considerable disaggregation available. In addition, the analytical sophistication implied by the technique generally requires adequate computer capabilities and a quantitatively oriented staff which can relate quantitative findings to persons not so inclined.

Systems analysis and simulation model

With the exception of Robin Barlow's pioneering work analyzing the economic impact of malaria eradication in Sri Lanka (Ceylon), I am not aware of any other attempts to develop and use a systems approach or to incorporate the use of a simulation model for analyzing the impact of major health programs on health and other socio-economic indicators [28]. While one can be critical of Barlow's efforts since he analyzed the impact of malaria eradication from the perspective of a single objective indicator (*per capita* income) without incorporating additional outcome indicators, particularly with respect to distributional impacts, his efforts must be applauded for operationalizing a macro-methodology appropriate for the set of primary, secondary and feedback effects which have occurred over time due to a major change in the disease structure of a population.

The comprehensiveness of the systems approach which manifests itself in simulation type computer-based models is both its major strength and weakness as an analytical tool for use in public policy analysis. In terms of its strength, it can be very useful in analyzing the time-paths of a number of interacting variables, as in the case of a disease eradication program, which can change not only the rate of growth of the population, but also the age structure, the productivity of the population and the pattern of demand for goods and services. Such a framework, assuming that it is properly specified, can also be used to monitor the impact of such a programmatic strategy on other outcome indicators, such as the distributional impact of the accruing benefits, specified either in health status or economic terms, as well as the government's ability to finance future programs demanded by the population. Perhaps one of the most important uses for which such a methodology could be employed at the present time would be an analysis of the health and other socio-economic effects of a significant increase in nutritional intake of not only calories and protein, but also other vitamin and mineral requirements, in developing countries. A number of health as well as economic policies would be simultaneously evaluated upon the development of such a model. In the case of the more developed countries, a model of the type described could be most appropriately used to analyze the effect of programs designed to significantly reduce the incidence of cardio-vascular diseases.

The comprehensiveness of the system's framework also retards its development and usefulness, however,

First, it is complex and requires considerable professional and computer time and effort to develop such a policy-oriented simulation model. The cost of developing such a model may be substantially mitigated; however, if it yields significantly improved information on a number of health programs under consideration [29]. Second, such models may not be correctly specified. Without being fully cognizant of the potential misspecification, many policies may be incorrectly analyzed. Third, often certain relationships cannot be adequately defined due to present knowledge and research constraints [30]. The results of the policy analysis thus rest upon the judgement and creativity of the analysts in ascertaining the appropriate relationships between the variables, using limited information. To a certain extent, this problem can be overcome by employing sensitivity analysis on certain key variables where information is lacking.

CONCLUDING COMMENTS

The review presented has focused on a number of problems in using benefit-cost analysis and other analytical tools as a means of assisting public decision-makers in allocating scarce resources to health. It is useful to conclude this analysis, however, by considering two constructive suggestions related to the efficacy of benefit-cost and systems analysis in planning and evaluating the use of resources in delivering health and medical care.

First, it is useful to reconsider the efficacy of using benefit-cost analysis in health program analysis. While it may not always be feasible to measure, with any degree of precision, the potential benefits and costs of a set of alternative health investments, the methodology is useful in providing a conceptual framework for at least enumerating the possible benefits and costs, in the form of a balance sheet, which can then be considered (at least in a qualitative way) in making programmatic decisions. Using the framework in this rudimentary fashion assists in systematizing the decision-making process and provides a basis for determining the priorities for acquiring additional information in order to improve the allocation process in the future. While costs and benefits are enumerated, concepts, goals and programmatic indicators can also be defined and staff personnel can begin to give systematic attention to problem areas.

Secondly, the systems framework can be helpful in expanding one's horizons relative to the feasible set of alternative health investments. While Fig. 1 has been developed for use in the African context, it provides a schematic example of the power of the systems analytical framework for analyzing, in a qualitative way, alternative investment opportunities normally considered external to the perview of health policy-makers. Investments in agricultural production of food processing, for example, may yield a greater health impact with fewer negative side effects in a particular country at a particular point in time than potential health investment alternatives in the health service sector. In addition, such a systems approach may provide the catalyst for more cooperative efforts between various Ministries which have jurisdiction over other social and economic sectors and the Ministry of Health, whose interest it is to promote the health of the country's people. As a consequence of such an interaction scheme, the Ministry of Health can better determine where it can provide assistance to other institutional structures which also have the capacity to improve the health of people.

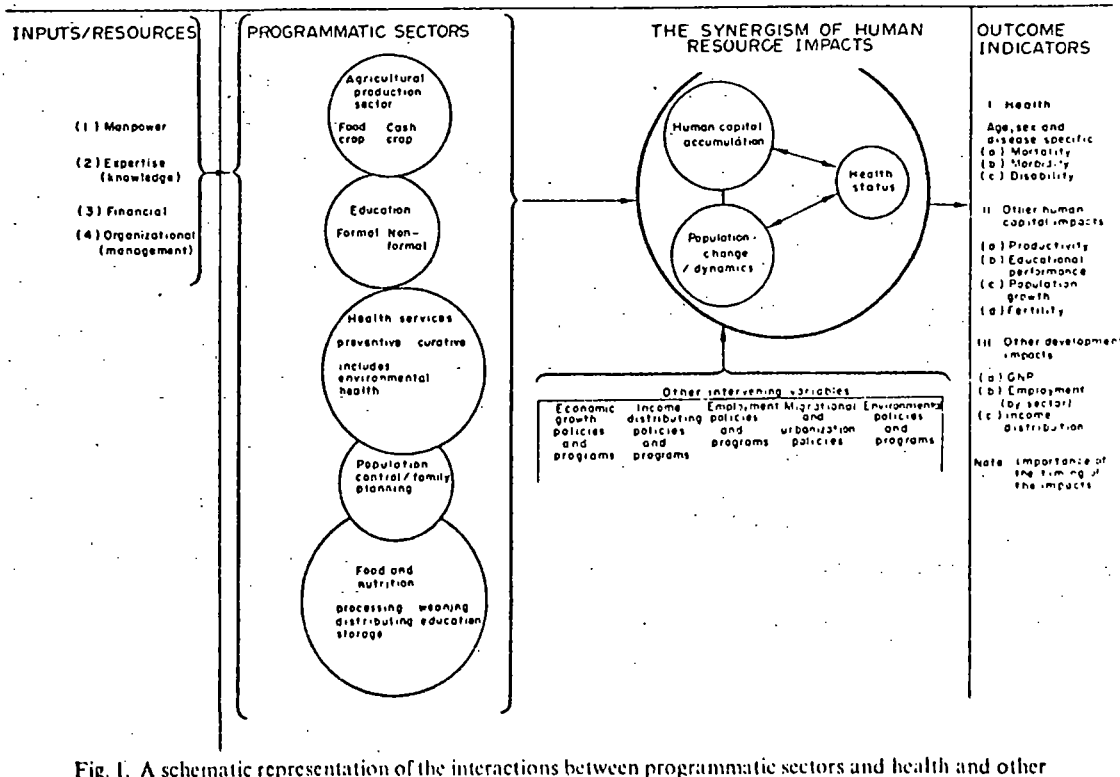


Fig. 1. A schematic representation of the interactions between programmatic sectors and health and other human resources impacts in developing countries (of Africa).

This review of alternative methods of evaluating health investments for use by public policy-makers is concluded by suggesting that one retain a pluralistic perspective. What is required is a sound analysis and statement of the problem to be investigated. Given a thorough understanding of the problem and the programmatic alternatives, an appropriate analytical methodology to assist in the decision-making process can be identified. At that point, the rate at which the analysis is conducted rests on the extent to which there exists information sufficient to understand the relationships between intervention and positive health outcomes.

REFERENCES

1. See the 3-vol. report of the Joint Economic Committee of the U.S. Congress for a thorough discussion of the initiation, development and evaluation of benefit-cost and related analytical frameworks in use by the U.S. Government. Joint Economic Committee, *The Analysis and Evaluation of Public Expenditures: The PPB System, A Compendium of Papers Submitted to the Sub-committee on Economy in Government*, 91st Congress, 1st Session, U.S. Government Printing, Washington, D.C., 1969. (3 Vols.). For particular guidance in the use of PPB in health in the U.S., see Robert N. Grosse, *Problems of Resource Allocation in Health, ibid.*, Vol. 3, pp. 1197–1223. For a continuing source of information on new developments in benefit-cost analysis, see the annual publication by Aldine-Atherton, begun 1971, entitled *Benefit-Cost Analysis 1971: An Aldine Annual*, Aldine-Atherton, Chicago, 1972.
2. E.g. see WHO, *National Health Planning in Developing Countries*, Tech. Report Series, 350, 1967; PAHO, *Health Planning–Problems of Concept and Method*, Scientific Publications No. 111, Washington, D.C., 1965 and the set of articles on Health and Socio-Economic Development, *Int. J. Hlth. Serv.*, 1, (3), 187, 1971.
3. See Kuhner, A. The Impact of Public Health Programs on Economic Development. Report of a study of malaria in Thailand. *Int. J. Hlth. Serv.*, 1, (3), 235, 1971; and Winslow, C. E. A. *The Cost of Sickness and the Price of Health*, WHO, Geneva, 1951.
4. Rice, D. *Economic Costs of Cardiovascular Diseases and Cancer, 1962*, Health Economics Series (5) U.S. Dept. of Health, Education and Welfare, Public Health Service, U.S. Government Printing Office, Washington, D.C., 1965. See also Klarman, H. Socio-economic Impact of Heart Disease. In: *2nd Nat. Conf. Cardiovascular Diseases, The Heart and Circulation*, 2, 631. U.S. Government Printing Office, Washington, D.C., 1965.
5. Weisbrod, B. *Economics of Public Health: Measuring the Economic Impact of Diseases*, University of Pennsylvania Press, Philadelphia, 1961. See also Weisbrod, B. Costs and Benefits of Medical Research: A Cost Study of Poliomyelitis. *Journal of Political Economy*, 79, 527, 1971.
6. Klarman, H. Syphilis Control Programs In: Dorfman, R. (editor) *Measuring Benefits of Government Investments* p. 367. Brookings Institution, Washington, D.C., p. 367.
7. U.S. Dept. of Health, Education and Welfare, Office of the Assistant Secretary (Planning and Evaluation), *Kidney Disease*, December, 1967.
8. Mushkin, S. Health as an Investment. *J. Pol. Econ.* 70, (5), 129, 1962.
9. U.S. Dept. of Health, Education and Welfare, Office of the Assistant Secretary for Program Coordination, *Motor Vehicle Injury Prevention Program*, August, 1966.
10. See Weisbrod, B. *Costs and Benefits of Medical Research: A Case Study of Poliomyelitis*, 1971. See also Crystal, R. and Brewster, A. Cost-Benefit and Cost-Effectiveness Analyses in the Health Field: An Introduction. *Inquiry*, 3, December, 1966.
11. See, e.g. Lindsay, C. Real Returns to Medical Education. *J. Human Resources*, 8, (3), 331, 1973. Hansen, W. L. Shortages and Investment in Health Manpower. In: *The Economics of Health and Medical Care*, p. 75. University of Michigan, Ann Arbor, 1964.
12. See Enke, S. The Economic Aspects of Solving Population Growth. *The Economic Journal*, 74, 301, 1966. For a comprehensive review of the population policy of the U.S., see Blumstein, J. The Development of U.S. Population Policy. *Milbank Memorial Fund Quarterly, Health and Society*, 51, 1973.
13. See Blayney, J. R. and Hill, I. N. Flourine and Dental Caries. *J. Am. Dent. Ass.* (Special Issue) January, 1967. For additional insight into the impact of flouridation, see also Douglas, B., Wallace, D., Lerner, M. and Coppersmith, S. Impact of Water Flouridation on Dental Practice and Dental Manpower. *J. Am. Dent. Ass.*, 84, 355, 1972.
14. The problem of not including qualitative indicators in evaluating the impact of education, a related human resource investment, has been widely discussed in the literature. See, e.g. Emmerij, L. *Can the School Build a New Social Order?* Elsevier Publishing Company, New York, 1974.
15. For a further discussion of this point, see Barlow, R. *The Economic Effects of Malaria Eradication*. Research Series No. 15, School of Public Health, University of Michigan, Ann Arbor, 1968.
16. See, e.g. Ruderman, A. P. Lessons from Latin American Experience, p. 13. U.S. Dept. of Health, Education and Welfare, *Economic Benefits from Public Health Services*. Public Health Service, U.S. Government Printing Office, Washington, D.C., 1964; Klarman, H. Present Status of Cost-Benefit Analysis in the Health Field. *Am. J. Publ. Hlth.*, 57, 11, 1967. Rado, E. The Benefit-Cost Analysis in Health. In: UNRISD. *The Use of Benefit-Cost Analysis in Social Development*. Report No. 5. UN Research Institute for Social Development, 1966.
17. For a theoretical discussion of this point, see Prest, A. R. and Turney, R. Cost-Benefit Analysis: A Survey. *The Economic Journal*, 75, 683, 1965.
18. See general books on project appraisal, e.g. Gittinger, J. P. *Economic Analysis of Agricultural Projects*, Johns Hopkins, Baltimore, 1972 and O.E.C.D., *Manual of Industrial Project Analysis in Developing Countries*, I, O.E.C.D., Paris, 1968.
19. For a further description of this methodology, see Nwaneri, V. C. Income Distribution and Project Selection. *Finance and Development*, 10, (3), 1973. See also Gershenberg, I. and Haskell, M. The Distribution of Medical Services in Uganda. *Social Science and Medicine*, 6, (3), 353, 1972, as an example of distributional analysis of health services in a developing country.
20. See Dunlop, D. W. The Development of an Output Concept for Analysis of Curative Health Services. *Social Science and Medicine*, 6, (3), 373, 1972; and Dunlop, D. W. *The Economics of Uganda's Health Service System: Implications for Health and Economic Planning*, unpublished Ph.D. Dissertation, Michigan State University, 1973. See also, Feldstein, M. S., Piot, M. A. and Sundaresan, T. K. Resource Allocation Model for Public Health Planning: A Case Study of Tuberculosis Control. *Bull. WHO* (Supplement) 1973.
21. To my knowledge, the U.S. is the only country in the world which regularly attempts to monitor the health status of its population using survey and clinical examination techniques.
22. The work of the National Center for Health Statistics and the Office of Research and Statistics, Social Secur-

ity Administration in the Department of Health, Education and Welfare are to be commended for their efforts made in this area. Work on improving international data is also to be commended.

23. See Joint Economic Committee, *The Analysis and Evaluation of Public Expenditures: The PPB System*, 1969, and Haveman, R. and Margolis, J. (Editors) *Public Expenditures and Policy Analysis*, Markham, Chicago, 1970.
24. Examples of the application of linear programming techniques in the health field include Feldstein, M. S. *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service*, North-Holland, Amsterdam, 1967 (Chapter 5); Feldstein, M. S. *et al.*, *Resource Allocation Model for Public Health Planning: A Case Study of Tuberculosis Control*, 1973; Dunlop, D. W. *The Economics of Uganda's Health Service System*, 1973, and Sheldon, A., Baker, F. and McLaughlin, C. (Editors) *Systems and Medical Care*, MIT Press, Cambridge, 1970.
25. For a more thorough discussion of these issues, see Dunlop, D. W. *The Economics of Uganda's Health Service System*, Chapters 4 and 5, 1973.
26. See Hillier, F. and Lieberman, G. *Introduction to Operations Research*, Holden Day, San Francisco, 1967 or Gass, S. *Linear Programming Methods and Applications*, McGraw-Hill, 1964, for a complete discussion of the problems and methodology of linear programming.
27. See Dunlop, D. W. The Development of an Output Concept for Analysis of Curative Health Services, 1972.
28. Barlow, R. *The Economic Effects of Malaria Eradication*, 1968.
29. A case in point in a related area are the simulation models developed for analyzing agricultural policies on a number of economic indicators for Korea and Nigeria. See Manetsch, T. S. *et al.*, *A Generalized Simulation Approach for Agricultural Sector Analysis with Special Reference to Nigeria*, Michigan State University, East Lansing, 1971.
30. The identification of knowledge gaps, however, may assist one in defining a more comprehensive set of priorities from which to develop future research activities.

Paul Lerman

r study director of Mobilization
:962.
Organization for Treatment:
New York: Free Press, 1966),
comment about "Inland," a pri-
mpleting treatment are actually
This study focuses on the atti-
residence. It would have been
s differed in their responses to
ole to know whether the runa-
d evidence that the attitudinal
ment" (i.e.; those who were
ent to questions about them-
e data are quite central to the
land.

s with Small vs. Large Living
Report No. 47 (Sacramento:
r 1, 1965). (Mimeographed.)

le is composed only of Negro
nted no "deep-seated psychiat-
ch more heterogeneous and has
be found in Lois Wiley, "An
ned master's thesis, New York

appraisals of the juvenile court
Matza, *Delinquency and Drift*

K. Turner, "Community Treat-
f Delinquents," CTP Research
Division of Research, August

18. *Reforms as Experiments*

Donald T. Campbell

The United States and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available. Our readiness for this stage is indicated by the inclusion of specific provisions for program evaluation in the first wave of the "Great Society" legislation, and by the current congressional proposals for establishing "social indicators" and socially relevant "data banks." So long have we had good intentions in this regard that many may feel we are already at this stage, that we already are continuing or discontinuing programs on the basis of assessed effectiveness. It is a theme of this article that this is not at all so, that most ameliorative programs end up with *no* interpretable evaluation (Etzioni, 1968; Hyman & Wright, 1967; Schwartz, 1961). We must look hard at the sources of this condition, and design ways of overcoming the difficulties. This article is a preliminary effort in this regard.

Many of the difficulties lie in the intransigencies of the research setting and in the presence of recurrent seductive pitfalls of interpretation. The bulk of this article will be devoted to these problems. But the few available solutions turn out to depend upon correct administrative decisions in the initiation and execution of the program. These decisions are made in a political arena, and involve political jeopardies that are often sufficient to explain the lack of hard-headed evaluation of effects. Removing reform administrators from the political spotlight seems both highly unlikely, and undesirable even if it were possible. What is instead essential is that the social scientist research advisor understand the political realities of the situation, and that he aid by helping create a public demand for hard-headed evaluation, by contributing to those political inventions that reduce the liability of honest evaluation, and by educating future administrators to the problems and possibilities.

For this reason, there is also an at-

The preparation of this paper has been supported by National Science Foundation Grant GS1309X. Versions of this paper have been presented at the Northwestern University Alumni Fund Lecture, January 24, 1968; to the Social Psychology Section of the British Psychological Society at Oxford, September 20, 1968; to the International Conference on Social Psychology at Prague, October 7, 1968 (under a different title); and to several other groups.

Reprinted with permission from *American Psychologist*, Vol. 24, No. 4, April 1969, pp. 409–429, with modifications.

tempt in this article to consider the political setting of program evaluation, and to offer suggestions as to political postures that might further a truly experimental approach to social reform. Although such considerations will be distributed as a minor theme throughout this article, it seems convenient to begin with some general points of this political nature.

POLITICAL VULNERABILITY FROM KNOWING OUTCOMES

It is one of the most characteristic aspects of the present situation that *specific reforms are advocated as though they were certain to be successful.* For this reason, knowing outcomes has immediate political implications. Given the inherent difficulty of making significant improvements by the means usually provided and given the discrepancy between promise and possibility, most administrators wisely prefer to limit the evaluations to those the outcomes of which they can control, particularly insofar as published outcomes or press releases are concerned. Ambiguity, lack of truly comparable comparison bases, and lack of concrete evidence all work to increase the administrator's control over what gets said, or at least to reduce the bite of criticism in the case of actual failure. There is safety under the cloak of ignorance. Over and above this tie-in of advocacy and administration, there is another source of vulnerability in that the facts relevant to experimental program evaluation are also available to argue the general efficiency and honesty of administrators. The public availability of such facts reduces the privacy and security of at least some administrators.

Even where there are ideological commitments to a hard-headed evaluation of organizational efficiency, or to a scientific organization of society, these two jeopardies lead to the failure to evaluate organizational experiments realistically. If the political and administrative system has committed itself in advance to the correctness and efficacy of its reforms, it cannot tolerate learning of failure. To be truly scientific we must be able to experiment. We must be able to advocate without that excess of commitment that blinds us to reality testing.

This predicament, abetted by public apathy and by deliberate corruption, may prove in the long run to permanently preclude a truly experimental approach to social amelioration. But our needs and our hopes for a better society demand we make the effort. There are a few signs of hope. In the United States we have been able to achieve cost-of-living and unemployment indices that, however imperfect, have embarrassed the administrations that published them. We are able to conduct censuses that reduce the number of representatives a state has in Congress. These are grounds for optimism, although the corrupt tardiness of state governments in following their own constitutions in revising legislative districts illustrates the problem.

One simple shift in political posture which would reduce the problem is the shift from the advocacy of a specific reform to the advocacy of the seriousness of the problem, and hence to the advocacy of persistence in alternative reform efforts should the first one fail. The political stance would become: "This is a serious problem. We propose to initiate Policy A on an experimental basis. If after five years there has been no significant improvement, we will shift to Policy B." By making explicit that a given problem solution was only one of several

that the administrator or party good conscience advocate, and l ready a plausible alternative, th istrator could afford honest eval outcomes. Negative results, a the first program, would not j his job, for his job would be after the problem until somet found that worked.

Coupled with this should be moratorium on ad hominum c research, that is, on research de evaluate specific administrato than alternative policies. If v about the invasion-of-privacy pr the data banks and social indi the future (e.g., Sawyer & S 1968), the touchiest point is th of administrators. If we thre the measurement system will sabotaged in the innumerable v sible. While this may soun pessimistic, the recurrent ane administrators attempting to unwanted research findings cor of its accuracy. But we should l evaluate those alternative polic given administrator has the c implementing.

FIELD EXPERIMENTS AND QUASI-EXPERIMENTAL DESIG

In efforts to extend the log oratory experimentation into th and into settings not fully exp an inventory of threats to exp validity has been assembled, in which some 15 or 20 experim quasi-experimental designs h evaluated (Campbell, 1957 Campbell & Stanley, 1963). In ent article only three or four de be examined, and therefore not validity threats will be releva will provide useful backgroun

onal efficiency, or to a scien-
ıization of society, these two
lead to the failure to evaluate
onal experiments realistically.
tical and administrative system
ıitted itself in advance to the
s and efficacy of its reforms, it
erate learning of failure. To be
ıtific we must be able to exper-
e must be able to advocate
ıat excess of commitment that
o reality testing.
redicament, abetted by public
ıd by deliberate corruption,
e in the long run to perma-
clude a truly experimental ap-
social amelioration. But our
our hopes for a better society
e make the effort. There are a
of hope. In the United States
een able to achieve cost-of-liv-
unemployment indices that,
imperfect, have embarrassed
istrations that published them.
le to conduct censuses that re-
number of representatives a
n Congress. These are grounds
sm, although the corrupt tardi-
ate governments in following
constitutions in revising leg-
stricts illustrates the problem.
mple shift in political posture
uld reduce the problem is the
the advocacy of a specific re-
ne advocacy of the seriousness
oblem, and hence to the advo-
ersistence in alternative reform
uld the first one fail. The po-
ıce would become: "This is a
oblem. We propose to initiate
on an experimental basis. If
vears there has been no signifi-
ovement, we will shift to Pol-
making explicit that a given
olution was only one of several

that the administrator or party could in good conscience advocate, and by having ready a plausible alternative, the administrator could afford honest evaluation of outcomes. Negative results, a failure of the first program, would not jeopardize his job, for his job would be to keep after the problem until something was found that worked.

Coupled with this should be a general moratorium on ad hominum evaluative research, that is, on research designed to evaluate specific administrators rather than alternative policies. If we worry about the invasion-of-privacy problem in the data banks and social indicators of the future (e.g., Sawyer & Schechter, 1968), the touchiest point is the privacy of administrators. If we threaten this, the measurement system will surely be sabotaged in the innumerable ways possible. While this may sound unduly pessimistic, the recurrent anecdotes of administrators attempting to squelch unwanted research findings convince me of its accuracy. But we should be able to evaluate those alternative policies that a given administrator has the option of implementing.

FIELD EXPERIMENTS AND QUASI-EXPERIMENTAL DESIGNS

In efforts to extend the logic of laboratory experimentation into the "field," and into settings not fully experimental, an inventory of threats to experimental validity has been assembled, in terms of which some 15 or 20 experimental and quasi-experimental designs have been evaluated (Campbell, 1957, 1963; Campbell & Stanley, 1963). In the present article only three or four designs will be examined, and therefore not all of the validity threats will be relevant, but it will provide useful background to look briefly at them all. Following are nine threats to internal validity.[1]

1. *History:* events, other than the experimental treatment, occurring between pretest and posttest and thus providing alternate explanations of effects.

2. *Maturation:* processes within the respondents or observed social units producing changes as a function of the passage of time per se, such as growth, fatigue, secular trends, etc.

3. *Instability:* unreliability of measures, fluctuations in sampling persons or components, autonomous instability of repeated or "equivalent" measures. (This is the only threat to which statistical tests of significance are relevant.)

4. *Testing:* the effect of taking a test upon the scores of a second testing. The effect of publication of a social indicator upon subsequent readings of that indicator.

5. *Instrumentation:* in which changes in the calibration of a measuring instrument or changes in the observers or scores used may produce changes in the obtained measurements.

6. *Regression artifacts:* pseudo-shifts occurring when persons or treatment units have been selected upon the basis of their extreme scores.

7. *Selection:* biases resulting from differential recruitment of comparison groups, producing different mean levels on the measure of effects.

8. *Experimental mortality:* the differential loss of respondents from comparison groups.

9. *Selection-maturation interaction:* selection biases resulting in differential rates of "maturation" or autonomous change.

If a change or difference occurs, these are rival explanations that could be used

to explain away an effect and thus to deny that in this specific experiment any genuine effect of the experimental treatment had been demonstrated. These are faults that true experiments avoid, primarily through the use of randomization and control groups. In the approach here advocated, this checklist is used to evaluate specific quasi-experimental designs. This is evaluation, not rejection, for it often turns out that for a specific design in a specific setting the threat is implausible, or that there are supplementary data that can help rule it out even where randomization is impossible. The general ethic, here advocated for public administrators as well as social scientists, is to use the very best method possible, aiming at "true experiments" with random control groups. But where randomized treatments are not possible, a self-critical use of quasi-experimental designs is advocated. We must do the best we can with what is available to us.

Our posture vis-à-vis perfectionist critics from laboratory experimentation is more militant than this: the only threats to validity that we will allow to invalidate an experiment are those that admit of the status of empirical laws more dependable and more plausible than the law involving the treatment. The mere possibility of some alternative explanation is not enough—it is only the *plausible* rival hypotheses that are invalidating. Vis-à-vis correlational studies, on the other hand, our stance is one of greater conservatism. For example, because of the specific methodological trap of regression artifacts, the sociological tradition of "ex post facto" designs (Chapin, 1947; Greenwood, 1945) is totally rejected (Campbell & Stanley, 1963, pp. 240–241; 1966, pp. 70–71).

Threats to external validity, which follow, cover the validity problems involved in interpreting experimental results, the threats to valid generalization of the results to other settings, to other versions of the treatment, or to other measures of the effect:[2]

1. *Interaction effects of testing:* the effect of a pretest in increasing or decreasing the respondent's sensitivity or responsiveness to the experimental variable, thus making the results obtained for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the experimental respondents were selected.

2. *Interaction of selection and experimental treatment:* unrepresentative responsiveness of the treated population.

3. *Reactive effects of experimental arrangements:* "artificiality"; conditions making the experimental setting atypical of conditions of regular application of the treatment: "Hawthorne effects."

4. *Multiple-treatment interference:* where multiple treatments are jointly applied, effects atypical of the separate application of the treatments.

5. *Irrelevant responsiveness of measures:* all measures are complex, and all include irrelevant components that may produce apparent effects.

6. *Irrelevant replicability of treatments:* treatments are complex, and replications of them may fail to include those components actually responsible for the effects.

These threats apply equally to true experiments and quasi-experiments. They are particularly relevant to applied experimentation. In the cumulative history of our methodology, this class of threats was first noted as a critique of true experiments involving prete & Goodman, 1939; Solor Such experiments provided for generalizing to other *pr* lations, but the reactions of populations to the treatmen be quite different. As a resu been an advocacy of true designs obviating the pre bell, 1957; Schanck & Goo Solomon, 1949) and a sea reactive measures (Webb Schwartz, & Sechrest, 1966)

These threats to validity a background against which cuss several research designs appropriate for evaluating grams of social amelioratio the "interrupted time-series "control series design," "re continuity design," and varic periments." The order is fr but generally available desig ones that require more ad foresight and determination.

INTERRUPTED TIME-SERIES

By and large, when a politi tiates a reform it is put across the board, with the tc ing affected. In this settin comparison base is the r-c vious years. The usual mode tion is a casual version of a quasi-experimental design, group pretest-posttest design.

A convenient illustration c the 1955 Connecticut crac speeding, which Sociologist F Ross and I have been anal methodological illustration (C Ross, 1968; Glass, 1968; Ros bell, 1968). After a record h fic fatalities in 1955, Governo Ribicoff instituted an unpr

:s to external validity, which fol-
'er the validity problems in-
: interpreting experimental re-
threats to valid generalization
sults to other settings, to other
of the treatment, or to other
of the effect:[2]

eraction effects of testing: the
a pretest in increasing or de-
the respondent's sensitivity or
eness to the experimental var-
us making the results obtained
retested population unrepresent-
the effects of the experimental
for the unpretested universe
ich the experimental respond-
selected.

eraction of selection and exper-
reatment: unrepresentative re-
ness of the treated population.

active effects of experimental ar-
nts: "artificiality"; conditions
he experimental setting atypical
ions of regular application of
ment: "Hawthorne effects."

ltiple-treatment interference:
ultiple treatments are jointly ap-
ects atypical of the separate ap-
of the treatments.

levant responsiveness of meas-
measures are complex, and all
rrelevant components that may
apparent effects.

levant replicability of treat-
reatments are complex, and rep-
of them may fail to include
mponents actually responsible
ffects.

threats apply equally to true ex-
s and quasi-experiments. They
cularly relevant to applied ex-
ation. In the cumulative history
ethodology, this class of threats
noted as a critique of true ex-

periments involving pretests (Schanck & Goodman, 1939; Solomon, 1949). Such experiments provided a sound basis for generalizing to other *pretested* populations, but the reactions of unpretested populations to the treatment might well be quite different. As a result, there has been an advocacy of true experimental designs obviating the pretest (Campbell, 1957; Schanck & Goodman, 1939; Solomon, 1949) and a search for nonreactive measures (Webb, Campbell, Schwartz, & Sechrest, 1966).

These threats to validity will serve as a background against which we will discuss several research designs particularly appropriate for evaluating specific programs of social amelioration. These are the "interrupted time-series design," the "control series design," "regression discontinuity design," and various "true experiments." The order is from a weak but generally available design to stronger ones that require more administrative foresight and determination.

INTERRUPTED TIME-SERIES DESIGN

By and large, when a political unit initiates a reform it is put into effect across the board, with the total unit being affected. In this setting the only comparison base is the record of previous years. The usual mode of utilization is a casual version of a very weak quasi-experimental design, the one-group pretest-posttest design.

A convenient illustration comes from the 1955 Connecticut crackdown on speeding, which Sociologist H. Laurence Ross and I have been analyzing as a methodological illustration (Campbell & Ross, 1968; Glass, 1968; Ross & Campbell, 1968). After a record high of traffic fatalities in 1955, Governor Abraham Ribicoff instituted an unprecedentedly severe crackdown on speeding. At the end of a year of such enforcement there had been but 284 traffic deaths as compared with 324 the year before. In announcing this the Governor stated, "With the saving of 40 lives in 1956, a reduction of 12.3% from the 1955 motor vehicle death toll, we can say that the program is definitely worthwhile." These results are graphed in Figure 1, with a deliberate effort to make them look impressive.

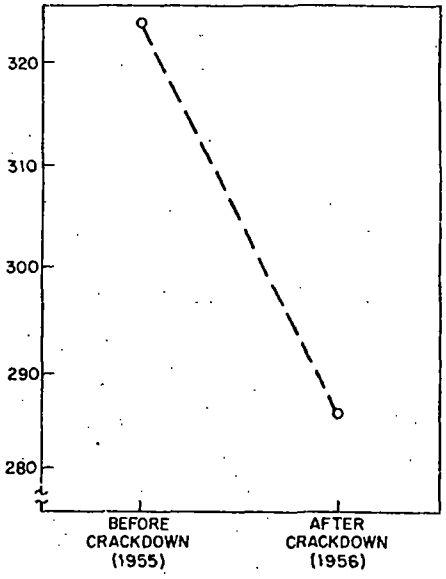


FIG. 1. Connecticut traffic fatalities.

In what follows, while we in the end decide that the crackdown had some beneficial effects, we criticize Ribicoff's interpretation of his results, from the point of view of the social scientist's proper standards of evidence. Were the now Senator Ribicoff not the man of stature that he is, this would be most unpolitic, because we could be alienating one of the strongest proponents of social experimentation in our nation.

Given his character, however, we may feel sure that he shares our interests both in a progressive program of experimental social amelioration, and in making the most hard-headed evaluation possible of these experiments. Indeed, it was his integrity in using every available means at his disposal as Governor to make sure that the unpopular speeding crackdown was indeed enforced that make these data worth examining at all. But the potentials of this one illustration and our political temptation to substitute for it a less touchy one, point to the political problems that must be faced in experimenting with social reform.

Keeping Figure 1 and Ribicoff's statement in mind, let us look at the same data presented as a part of an extended time series in Figure 2 and go over the relevant threats to internal validity. First, *History*. Both presentations fail to control for the effects of other potential change agents. For instance, 1956 might have been a particularly dry year, with fewer accidents due to rain or snow. Or there might have been a dramatic increase in use of seat belts, or other safety features. The advocated strategy in quasi-experimentation is not to throw up one's hands and refuse to use the evidence because of this lack of control, but rather to generate by informed criticism appropriate to this specific setting as many *plausible* rival hypotheses as possible, and then to do the supplementary research, as into weather records and safety-belt sales, for example, which would reflect on these rival hypotheses.

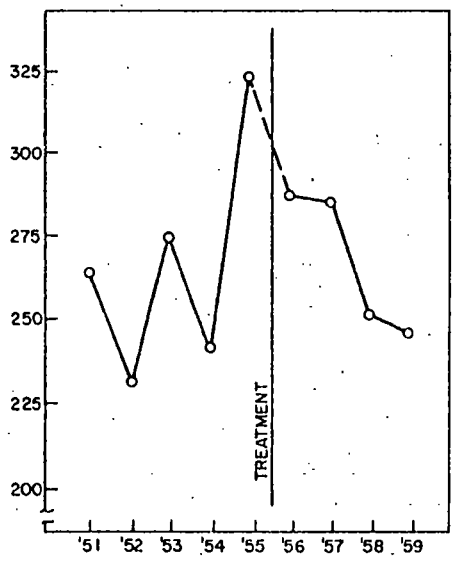


FIG. 2. Connecticut traffic fatalities. (Same data as in Figure 1 presented as part of an extended time series.)

Maturation. This is a term coming from criticisms of training studies of children. Applied here to the simple pretest-posttest data of Figure 1, it could be the plausible rival hypothesis that death rates were steadily going down year after year (as indeed they are, relative to miles driven or population of automobiles). Here the extended time series has a strong methodological advantage, and rules out this threat to validity. The general trend is inconsistently up prior to the crackdown, and steadily down thereafter.

Instability. Seemingly implicit in the public pronouncement was the assumption that all of the change from 1955 to 1956 was due to the crackdown. There was no recognition of the fact that all time series are unstable even when no treatments are being applied. The degree of this normal instability is the crucial issue, and one of the main advantages of the extended time series is that it samples this instability. The great pretreatment instability now makes the treatment effect look relatively trivial. The 1955–56 shift is less than the gains of both 1954–55 and 1952–53. It is the

largest drop in the series, b
the drops of 1951–52, 19
1957–58 by trivial amount
unexplained instabilities of t
such as to make the 1955–
derstandable as more of the s
other hand, it is noteworth
the crackdown there are no
gains, and in this respect t
of the time series seems
have changed.

The threat of instability
threat to which tests of sig
relevant. Box and Tiao (19
elegant Bayesian model fo
rupted time series. Applie
(1968) to our monthly dat
sonal trends removed, it sho
cally significant downward
series after the crackdown.
shall see, an alternative expla
least part of this significant e

Regression. In true expe
treatment is applied indep
the prior state of the units. It
periments exposure to treatm
a cosymptom of the treated
dition. The treatment is apt
fect rather than, or in additi
a cause. Psychotherapy is suc
tom treatment, as is any oth
the treated group is self-sel
signed on the basis of nee
present special problems of
tion, of which the present
provides one type.

The selection-regression
val hypothesis works this
that the fatality rate has son
unreliability, then a subsam
for its extremity in 1955 w
average, merely as a reflectio
reliability, be less extreme i
there been selection for extr
plying this treatment? Prob

ents. For instance, 1956 might
a particularly dry year, with
dents due to rain or snow. Or
ht have been a dramatic in-
se of seat belts, or other safety
The advocated strategy in
rimentation is not to throw
hands and refuse to use the
ecause of this lack of control,
to generate by informed crit-
opriate to this specific setting
ble rival hypotheses as
nd then to do the supplemen-
ch, as into weather records and
sales, for example, which
ect on these rival hypotheses.
ion. This is a term coming
cisms of training studies of
pplied here to the simple pre-
st data of Figure 1, it could be
ble rival hypothesis that death
steadily going down year after
indeed they are, relative to
en or population of automo-
re the extended time series
ng methodological advantage,
out this threat to validity. The
nd is inconsistently up prior
ackdown, and steadily down

ity. Seemingly implicit in the
nouncement was the assump-
ll of the change from 1955 to
due to the crackdown. There
cognition of the fact that all
s are unstable even when no
are being applied. The degree
rmal instability is the crucial
one of the main advantages of
ed time series is that it sam-
nstability. The great pretreat-
ibility now makes the treat-
t look relatively trivial. The
hift is less than the gains of
-55 and 1952–53. It is the largest drop in the series, but it exceeds the drops of 1951–52, 1953–54, and 1957–58 by trivial amounts. Thus the unexplained instabilities of the series are such as to make the 1955–56 drop understandable as more of the same. On the other hand, it is noteworthy that after the crackdown there are no year-to-year gains, and in this respect the character of the time series seems definitely to have changed.

The threat of instability is the only threat to which tests of significance are relevant. Box and Tiao (1965) have an elegant Bayesian model for the interrupted time series. Applied by Glass (1968) to our monthly data, with seasonal trends removed, it shows a statistically significant downward shift in the series after the crackdown. But as we shall see, an alternative explanation of at least part of this significant effect exists.

Regression. In true experiments the treatment is applied independently of the prior state of the units. In natural experiments exposure to treatment is often a cosymptom of the treated group's condition. The treatment is apt to be an *effect* rather than, or in addition to being, a cause. Psychotherapy is such a cosymptom treatment, as is any other in which the treated group is self-selected or assigned on the basis of need. These all present special problems of interpretation, of which the present illustration provides one type.

The selection-regression plausible rival hypothesis works this way: Given that the fatality rate has some degree of unreliability, then a subsample selected for its extremity in 1955 would on the average, merely as a reflection of that unreliability, be less extreme in 1956. Has there been selection for extremity in applying this treatment? Probably yes. Of all Connecticut fatality years, the most likely time for a crackdown would be after an exceptionally high year. If the time series showed instability, the subsequent year would on the average be less, *purely as a function of that instability.* Regression artifacts are probably the most recurrent form of self-deception in the experimental social reform literature. It is hard to make them intuitively obvious. Let us try again. Take any time series with variability, including one generated of pure error. Move along it as in a time dimension. Pick a point that is the "highest so far." Look then at the next point. On the average this next point will be lower, or nearer the general trend.

In our present setting the most striking shift in the whole series is the upward shift just prior to the crackdown. It is highly probable that this caused the crackdown, rather than, or in addition to, the crackdown causing the 1956 drop. At least part of the 1956 drop is an artifact of the 1955 extremity. While in principle the degree of expected regression can be computed from the autocorrelation of the series, we lack here an extended-enough body of data to do this with any confidence.

Advice to administrators who want to do genuine reality-testing must include attention to this problem, and it will be a very hard problem to surmount. The most general advice would be to work on chronic problems of a persistent urgency or extremity, rather than reacting to momentary extremes. The administrator should look at the pretreatment time series to judge whether or not instability plus momentary extremity will explain away his program gains. If it will, he should schedule the treatment for a year or two later, so that his deci-

sion is more independent of the one year's extremity. (The selection biases remaining under such a procedure need further examination.)

In giving advice to the *experimental* administrator, one is also inevitably giving advice to those *trapped* administrators whose political predicament requires a favorable outcome whether valid or not. To such trapped administrators the advice is pick the very worst year, and the very worst social unit. If there is inherent instability, there is no where to go but up, for the average case at least.

Two other threats to internal validity need discussion in regard to this design. By *testing* we typically have in mind the condition under which a test of attitude, ability, or personality is itself a change agent, persuading, informing, practicing, or otherwise setting processes of change in action. No artificially introduced testing procedures are involved here. However, for the simple before-and-after design of Figure 1, if the pretest were the first data collection of its kind ever publicized, this publicity in itself might produce a reduction in traffic deaths which would have taken place even without a speeding crackdown. Many traffic safety programs assume this. The longer time-series evidence reassures us on this only to the extent that we can assume that the figures had been published each year with equivalent emphasis.[3]

Instrumentation changes are not a likely flaw in this instance, but would be if recording practices and institutional responsibility had shifted simultaneously with the crackdown. Probably in a case like this it is better to use raw frequencies rather than indices whose correction parameters are subject to periodic revision. Thus per capita rates are subject to periodic jumps as new census figures become available correcting old extrapolations. Analogously, a change in the miles per gallon assumed in estimating traffic mileage for mileage-based mortality rates might explain a shift. Such biases can of course work to disguise a true effect. Almost certainly, Ribicoff's crackdown reduced traffic speed (Campbell & Ross, 1968). Such a decrease in speed increases the miles per gallon actually obtained, producing a concomitant drop in the estimate of miles driven, which would appear as an inflation of the estimate of mileage-based traffic fatalities if the same fixed approximation to actual miles per gallon were used, as it undoubtedly would be.

The "new broom" that introduces abrupt changes of policy is apt to reform the record keeping too, and thus confound reform treatments with instrumentation change. The ideal experimental administrator will, if possible, avoid doing this. He will prefer to keep comparable a partially imperfect measuring system rather than lose comparability altogether. The politics of the situation do not always make this possible, however. Consider, as an experimental reform, Orlando Wilson's reorganization of the police system in Chicago. Figure 3 shows his impact on petty larceny in Chicago—a striking *increase!* Wilson, of course, called this shot in advance, one aspect of his reform being a reform in the bookkeeping. (Note in the pre-Wilson records the suspicious absence of the expected upward secular trend.) In this situation Wilson had no choice. Had he left the record keeping as it was, for the purposes of better experimental design, his resentful patrolmen would have clobbered him with a crime wave by delib-

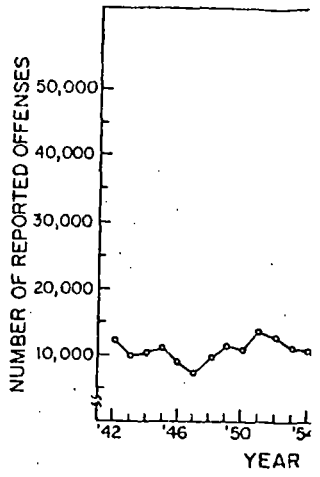


FIG. 3. Number of repo
under $50 in Chicago, Illino
to 1962 (data from *Unifor
ports for the United States*, 1

erately starting to record th
plaints that had not been
the books.[4]

Those who advocate the
val measures as social indica
1966; Gross, 1966, 1967; K
Webb et al., 1966) must
only to their high degree o
ror and systematic bias, but
politically motivated change
keeping that will follow upo
lic use as social indicators
Lehman, 1967). Not all m
equally susceptible. In Figure
Wilson's effect on homicides
ligible one way or the other.

Of the threats to external
one most relevant to social e
tion is *Irrelevant Respons
Measures*. This seems best d
terms of the problem of g
from indicator to indicator c
of the imperfect validity of
ures that is only to be overc
use of multiple measures of ir

mps as new census figures be-
able correcting old extrapola-
ogously, a change in the miles
assumed in estimating traffic
mileage-based mortality rates
ain a shift. Such biases can of
k to disguise a true effect. Al-
nly, Ribicoff's crackdown re-
ic speed (Campbell & Ross,
ch a decrease in speed in-
miles per gallon actually ob-
oducing a concomitant drop
mate of miles driven, which
ear as an inflation of the esti-
leage-based traffic fatalities if
ixed approximation to actual
gallon were used, as it un-
would be.
ew broom" that introduces
ages of policy is apt to reform
keeping too, and thus con-
rm treatments with instru-
change. The ideal experi-
ninistrator will, if possible,
g this. He will prefer to keep
a partially imperfect meas-
m rather than lose compara-
gether. The politics of the
o not always make this pos-
ver. Consider, as an experi-
orm, Orlando Wilson's reor-
of the police system in Chi-
re 3 shows his impact on
ıy in Chicago—a striking *in-*
ilson, of course, called this
vance, one aspect of his re-
; a reform in the bookkeep-
in the pre-Wilson records
ous absence of the expected
ular trend.) In this situation
d no choice. Had he left the
oing as it was, for the pur-
tter experimental design, his
atrolmen would have clob-
with a crime wave by delib-

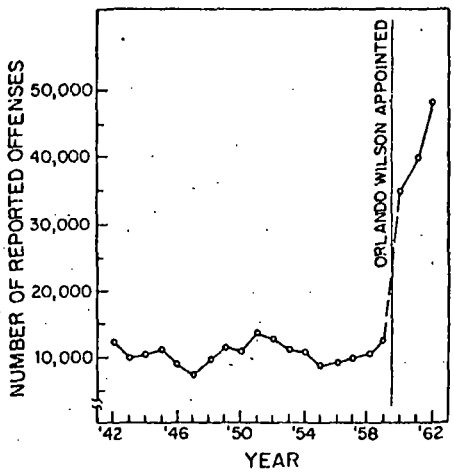


FIG. 3. Number of reported larcenies under $50 in Chicago, Illinois, from 1942 to 1962 (data from *Uniform Crime Reports for the United States*, 1942–62).

erately starting to record the many complaints that had not been getting into the books.[4]

Those who advocate the use of archival measures as social indicators (Bauer, 1966; Gross, 1966, 1967; Kaysen, 1967; Webb et al., 1966) must face up not only to their high degree of chaotic error and systematic bias, but also to the politically motivated changes in record keeping that will follow upon their public use as social indicators (Etzioni & Lehman, 1967). Not all measures are equally susceptible. In Figure 4, Orlando Wilson's effect on homicides seems negligible one way or the other.

Of the threats to external validity, the one most relevant to social experimentation is *Irrelevant Responsiveness of Measures*. This seems best discussed in terms of the problem of generalizing from indicator to indicator or in terms of the imperfect validity of all measures that is only to be overcome by the use of multiple measures of independent imperfection (Campbell & Fiske, 1959; Webb et al., 1966).

For treatments on any given problem within any given governmental or business subunit, there will usually be something of a governmental monopoly on reform. Even though different divisions may optimally be trying different reforms, within each division there will usually be only one reform on a given problem going on at a time. But for measures of effect this need not and should not be the case. The administrative machinery should itself make multiple measures of potential benefits and of unwanted side effects. In addition, the loyal opposition should be allowed to add still other indicators, with the political process and adversary argument challenging both validity and relative importance, with social science methodologists testifying for both parties, and with the basic records kept public and under bipartisan audit (as are voting

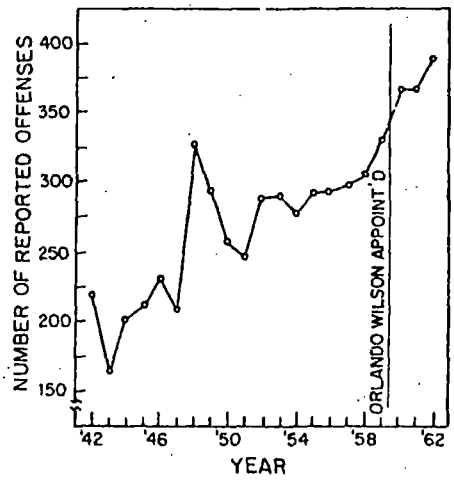


FIG. 4. Number of reported murders and nonnegligent manslaughters in Chicago, Illinois, from 1942 to 1962 (data from *Uniform Crime Reports for the United States*, 1942–62).

records under optimal conditions). This competitive scrutiny is indeed the main source of objectivity in sciences (Polanyi, 1966, 1967; Popper, 1963) and epitomizes an ideal of democratic practice in both judicial and legislative procedures.

The next few figures return again to the Connecticut crackdown on speeding and look to some other measures of effect. They are relevant to the confirming that there was indeed a crackdown, and to the issue of side effects. They also provide the methodological comfort of assuring us that in some cases the interrupted time-series design can provide clear-cut evidence of effect. Figure 5 shows the jump in suspensions of licen-

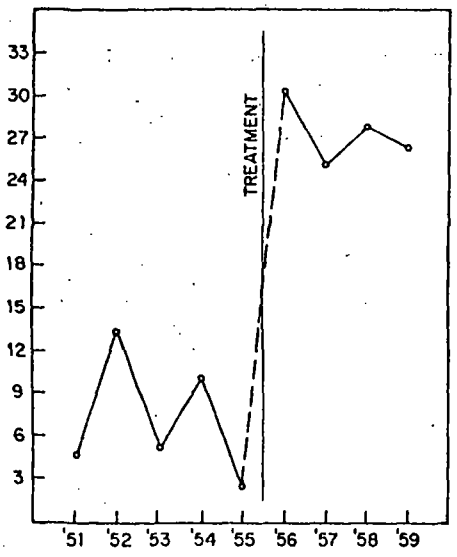


FIG. 5. Suspensions of licenses for speeding, as a percentage of all suspensions.

ses for speeding—evidence that severe punishment was abruptly instituted. Again a note to experimental administrators: with this weak design, *it is only abrupt and decisive changes that we have any chance of evaluating*. A gradually introduced reform will be indistinguishable from the background of secular change, from the net effect of the innumerable change agents continually impinging.

We would want intermediate evidence that traffic speed was modified. A sampling each year of a few hundred five-minute highway movies (random as to location and time) could have provided this at a moderate cost, but they were not collected. Of the public records available, perhaps the data of Figure 6, showing a reduction in speeding violations, indicate a reduction in traffic speed. But the effects on the legal system

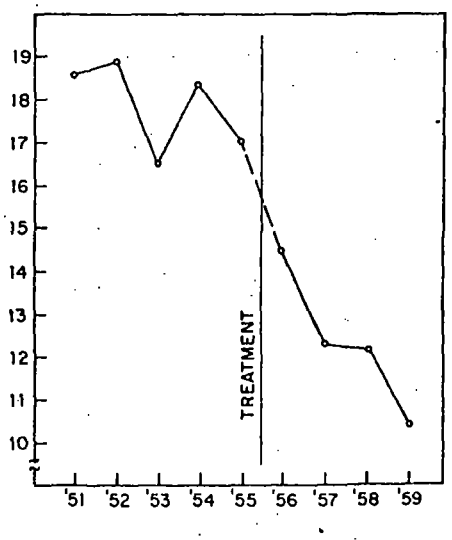


FIG. 6. Speeding violations, as a percentage of all traffic violations.

were complex, and in part undesirable. Driving with a suspended license markedly increased (Figure 7), at least in the biased sample of those arrested. Presumably because of the harshness of the punishment if guilty, judges may

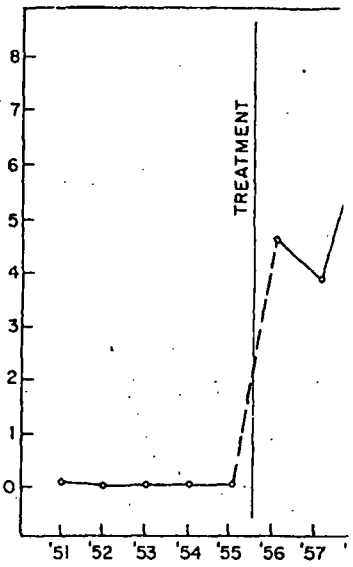


FIG. 7. Arrested while drivin suspended license, as a percentag pensions.

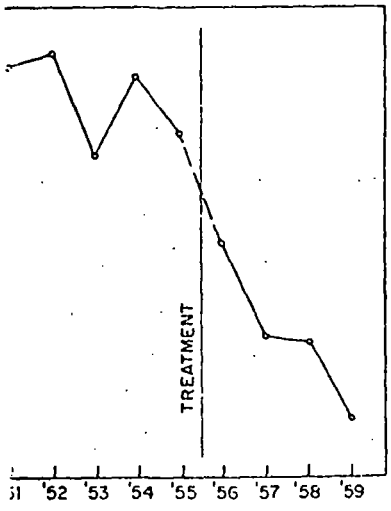
have become more lenient (Fi although this effect is of marg nificance.

The relevance of indicators fo cial problems we wish to cure kept continually in focus. The s dicators approach will tend tc n indicators themselves the goal c action, rather than the social p they but imperfectly indicate. Tl apt to be tendencies to legislate in the indicators per se rath changes in the social problems.

To illustrate the problem of tl evant responsiveness of measur ure 9 shows a result of the 1900 in divorce law in Germany. In a reanalysis of the data with the B Tiao (1965) statistic, Glass Tiao, & Maguire, 1969) has fou change highly significant, in con earlier statistical analyses (Rhe 1959; Wolf, Lüke, & Hax, 1959

t and decisive changes that we have bance of evaluating. A gradually in- ced reform will be indistinguish- from the background of secular e, from the net effect of the innu- ble change agents continually im- ng.

e would want intermediate evi- that traffic speed was modified. A ing each year of a few hundred inute highway movies (random as cation and time) could have pro- this at a moderate cost, but they not collected. Of the public records ible, perhaps the data of Figure 6, ng a reduction in speeding viola- indicate a reduction in traffic . But the effects on the legal system



6. Speeding violations, as a per- e of all traffic violations.

complex, and in part undesirable. ig with a suspended license mark- ncreased (Figure 7), at least in iased sample of those arrested. nably because of the harshness of nishment if guilty, judges may

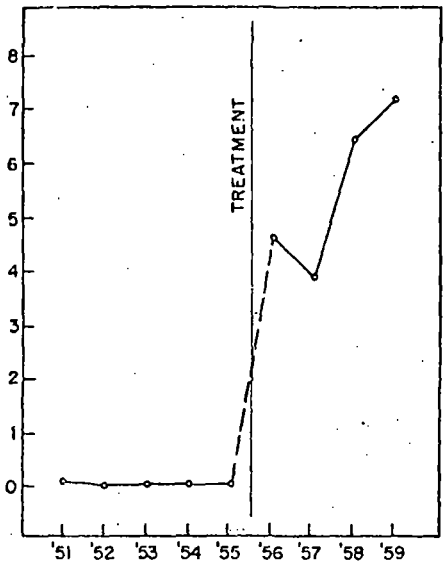


FIG. 7. Arrested while driving with a suspended license, as a percentage of suspensions.

have become more lenient (Figure 8) although this effect is of marginal significance.

The relevance of indicators for the social problems we wish to cure must be kept continually in focus. The social indicators approach will tend to make the indicators themselves the goal of social action, rather than the social problems they but imperfectly indicate. There are apt to be tendencies to legislate changes in the indicators *per se* rather than changes in the social problems.

To illustrate the problem of the irrelevant responsiveness of measures, Figure 9 shows a result of the 1900 change in divorce law in Germany. In a recent reanalysis of the data with the Box and Tiao (1965) statistic, Glass (Glass, Tiao, & Maguire, 1969) has found the change highly significant, in contrast to earlier statistical analyses (Rheinstein, 1959; Wolf, Lüke, & Hax, 1959). But

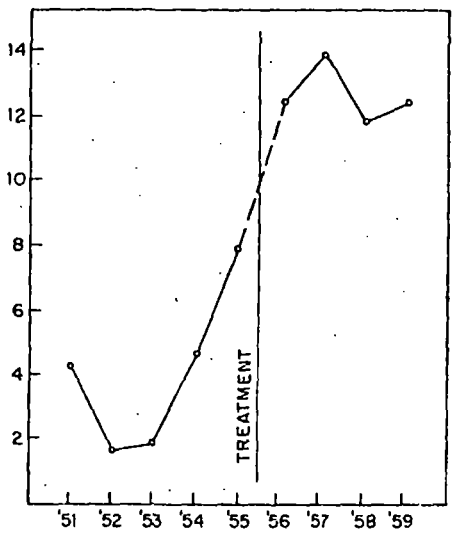


FIG. 8. Percentage of speeding violations judged not guilty.

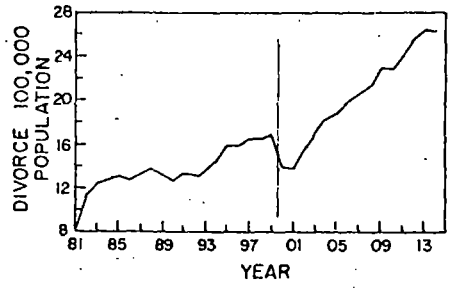


FIG. 9. Divorce rate for German Empire, 1881–1914.

Rheinstein's emphasis would still be relevant: This indicator change indicates no likely improvement in marital harmony, or even in marital stability. Rather than reducing them, the legal change has made the divorce rate a less valid indicator of marital discord and separation than it had been earlier (see also Etzioni & Lehman, 1967).

CONTROL SERIES DESIGN

The interrupted time-series design as discussed so far is available for those settings in which no control group is possi-

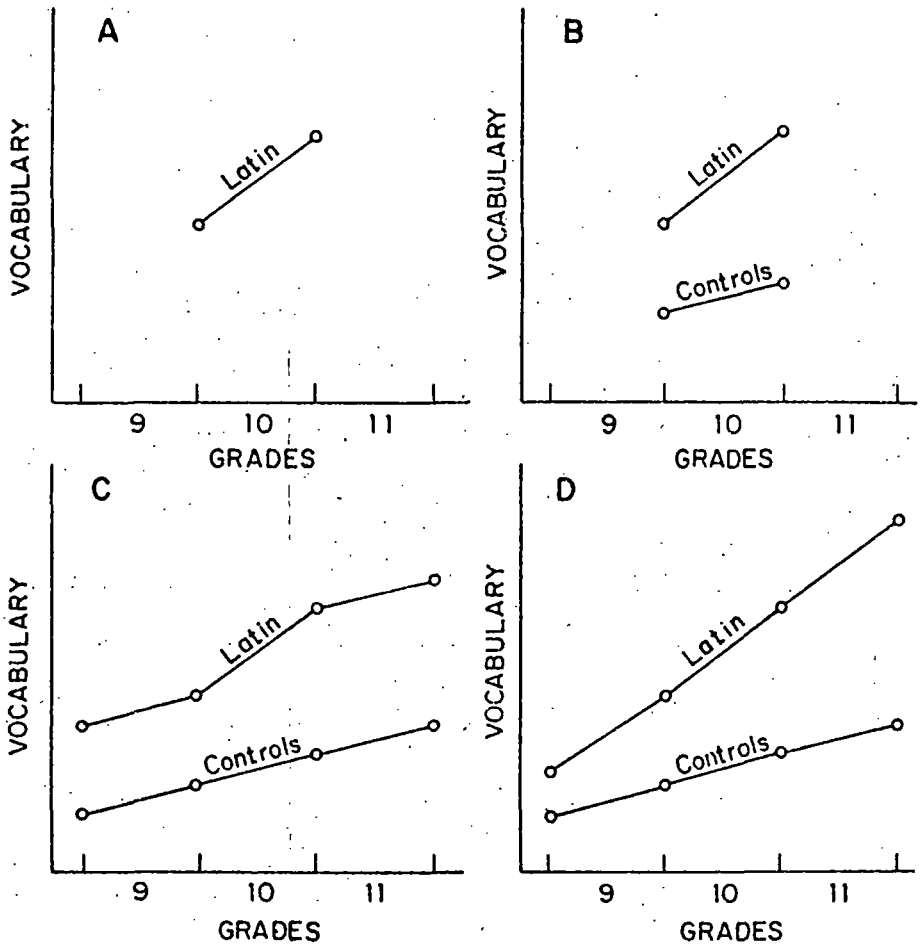


FIG. 10. Forms of quasi-experimental analysis for the effect of specific course work, including control series design.

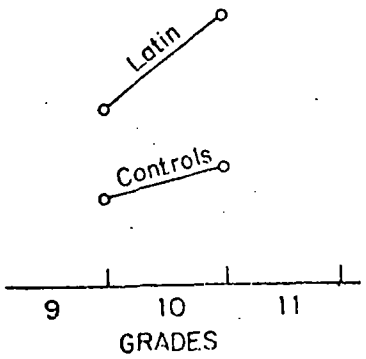
ble, in which the total governmental unit has received the experimental treatment, the social reform measure. In the general program of quasi-experimental design, we argue the great advantage of untreated comparison groups even where these cannot be assigned at random. The most common of such designs is the nonequivalent control-group pretest-posttest design, in which for each of two natural groups, one of which receives the treatment, a pretest and posttest measure is taken. If the traditional mistaken practice is avoided of matching on pretest scores (with resultant regression artifacts), this design provides a useful control over those aspects of history, maturation, and test-retest effects shared by both groups. But it does not control for the plausible rival hypothe-

sis of *selection-maturation in*
that is, the hypothesis that th
differences in the natural ag
involve not only differences
level, but differences in matura

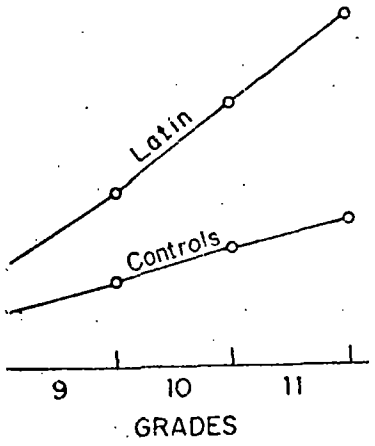
This point can be illustrate
of the traditional quasi-experi
sign problem of the effects o
English vocabulary (Campbe
In the hypothetical data of Fi
two alternative interpretatio
open. Latin may have had
those taking Latin gained r
those not. But, on the other h
students taking Latin may hav
annual rate of vocabulary gr
would manifest itself wheth
they took Latin. Extending thi
design into two time series pro
vant evidence, as comparison
alternative outcomes of Figur
10D shows. Thus approaching
perimental design from eithe
ing the nonequivalent control-
sign or from improving the i
time-series design, we arrive a
trol series design. Figure 11 s
for the Connecticut speeding c
adding evidence from the fat
of neighboring states. Here th
presented as population-base
rates so as to make the two seri
parable magnitude.

The control series design of
shows that downward trends w
able in the other states for 19
due to history and maturatio
due to shared secular trends,
automotive safety features, etc
data also show a general trend
necticut to rise relatively clos
other states prior to 1955, and
ily drop more rapidly than ot
from 1956 on. Glass (1968)
our monthly data for Connec

3



D



the effect of specific course work,

itural groups, one of which re-
the treatment, a pretest and post-
easure is taken. If the traditional
en practice is avoided of match-
pretest scores (with resultant re-
n artifacts), this design provides
l control over those aspects of his-
naturation, and test-retest effects
by both groups. But it does not
for the plausible rival hypothe-

sis of *selection-maturation interaction*—that is, the hypothesis that the selection differences in the natural aggregations involve not only differences in mean level, but differences in maturation rate.

This point can be illustrated in terms of the traditional quasi-experimental design problem of the effects of Latin on English vocabulary (Campbell, 1963). In the hypothetical data of Figure 10B, two alternative interpretations remain open. Latin may have had effect, for those taking Latin gained more than those not. But, on the other hand, those students taking Latin may have a greater annual rate of vocabulary growth that would manifest itself whether or not they took Latin. Extending this common design into two time series provides relevant evidence, as comparison of the two alternative outcomes of Figure 10C and 10D shows. Thus approaching quasi-experimental design from either improving the nonequivalent control-group design or from improving the interrupted time-series design, we arrive at the control series design. Figure 11 shows this for the Connecticut speeding crackdown, adding evidence from the fatality rates of neighboring states. Here the data are presented as population-based fatality rates so as to make the two series of comparable magnitude.

The control series design of Figure 11 shows that downward trends were available in the other states for 1955–56 as due to history and maturation, that is, due to shared secular trends, weather, automotive safety features, etc. But the data also show a general trend for Connecticut to rise relatively closer to the other states prior to 1955, and to steadily drop more rapidly than other states from 1956 on. Glass (1968) has used our monthly data for Connecticut and

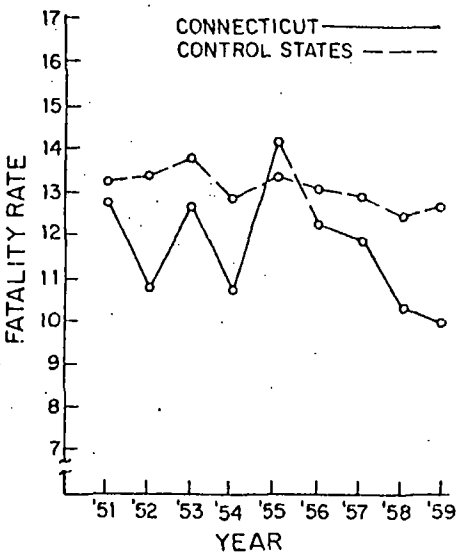


FIG. 11. Control series design comparing Connecticut fatalities with those of four comparable states.

the control states to generate a monthly difference score, and this too shows a significant shift in trend in the Box and Tiao (1965) statistic. Impressed particularly by the 1957, 1958, and 1959 trend, we are willing to conclude that the crackdown had some effect, over and above the undeniable pseudo-effects of regression (Campbell & Ross, 1968).

The advantages of the control series design point to the advantages for social experimentation of a social system allowing subunit diversity. Our ability to estimate the effects of the speeding crackdown, Rose's (1952) and Stieber's (1949) ability to estimate the effects on strikes of compulsory arbitration laws, and Simon's (1966) ability to estimate the price elasticity of liquor were made possible because the changes were not being put into effect in all states simultaneously, because they were matters of state legislation rather than national. I

do not want to appear to justify on these grounds the wasteful and unjust diversity of laws and enforcement practices from state to state. But I would strongly advocate that social engineers make use of this diversity while it remains available, and plan cooperatively their changes in administrative policy and in record keeping so as to provide optimal experimental inference. More important is the recommendation that, for those aspects of social reform handled by the central government, a purposeful diversity of implementation be envisaged so that experimental and control groups be available for analysis. Properly planned, these can approach true experiments, better than the casual and ad hoc comparison groups now available. But without such fundamental planning, uniform central control can reduce the present possibilities of reality testing, that is, of true social experimentation. In the same spirit, decentralization of decision making, both within large government and within private monopolies, can provide a useful competition for both efficiency and innovation, reflected in a multiplicity of indicators.

One further illustration of the interrupted time series and the control series will be provided. The variety of illustrations so far given have each illustrated some methodological point, and have thus ended up as "bad examples." To provide a "good example," an instance which survives methodological critique as a valid illustration of a successful reform, data from the British Road Safety Act of 1967 are provided in Figure 11A (from Ross, Campbell, & Glass, 1970).

The data on a weekly-hours basis are available only for a composite category of fatalities plus serious injuries, and Figure 11A therefore uses this composite for all three bodies of data. The "Weekend Nights" comprises Friday and Saturday nights from 10:00 P.M. to 4:00 A.M. Here, as expected, the crackdown is most dramatically effective, producing initially more than a 40 per cent drop, leveling off at perhaps 30 per cent, although this involves dubious extrapolations in the absence of some control comparison to indicate what the trend over the years might have been without the crackdown. In this British case, no comparison state with comparable traffic conditions or drinking laws was available. But controls need not always be separate groups of persons, they may also be separate samples of times or stimulus materials (Campbell & Stanley, 1966 pp. 43–47).

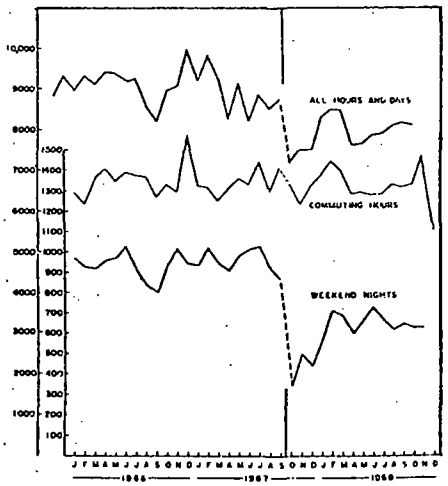


FIG. 11A. British traffic fatalities plus serious injuries, before and after Breathalyser crackdown of October 1967 (seasonally adjusted).

A cigarette company may use the sales of its main competitor as a control comparison to evaluate a new advertising campaign. One should search around for

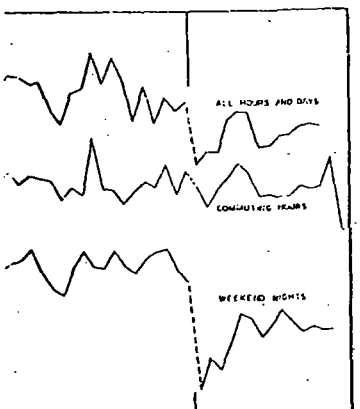
the most nearly appropriate co
parison. For the Breathalyser c
commuting hours when pubs
long closed seemed ideal. (T
muting Hours" figures come
A.M. to 10:00 A.M. and 4:(
5:00 P.M. Monday through Fr
are open for lunch from 12:(
or 2:30, and open again at 5:0

These commuting hours
vincingly show no effect, but
stable to help much with esti
long-term effects. They show
annual cycle than do the week
or the overall figures, and do n
far enough to provide an ade
for estimating this annual
precision.

The use of a highly judgm
gory such as "serious injuries
an opportunity for pseudo eff
to a shift in the classifiers' stan
overall figures are available
for fatalities, and these show
significant effect as strong as
for the serious injury categc
composite shown in Figure 11.

More details and the meth
problems are considered in
presentation (Ross, Campbell
1970). One further rule for
this design needs emphasi
terrupted time series can pro
evidence of effect only where
is introduced with a vigorous a
A gradually introduced reform
chance of being distinguis
shifts in secular trends or
cumulative effect of the many
fluences impinging during a
period of introduction. In th
alyser crackdown, an intense
campaign naming the specifi
date preceded the actual crack

herefore uses this composite for
bodies of data. The "Weekend
comprises Friday and Saturday
om 10:00 P.M. to 4:00 A.M.
xpected, the crackdown is most
lly effective, producing ini-
re than a 40 per cent drop, lev-
it perhaps 30 per cent, although
ives dubious extrapolations in
ce of some control comparison
e what the trend over the years
ve been without the crackdown.
itish case, no comparison state
nparable traffic conditions or
laws was available. But controls
always be separate groups of
they may also be separate sam-
times or stimulus materials
ll & Stanley, 1966 pp. 43–47).



A. British traffic fatalities plus
juries, before and after Breath-
kdown of October 1967 (season-
ted).

tte company may use the sales of
competitor as a control com-
to evaluate a new advertising
n. One should search around for

the most nearly appropriate control comparison. For the Breathalyser crackdown, commuting hours when pubs had been long closed seemed ideal. (The "Commuting Hours" figures come from 7:00 A.M. to 10:00 A.M. and 4:00 P.M. to 5:00 P.M. Monday through Friday. Pubs are open for lunch from 12:00 to 2:00 or 2:30, and open again at 5:00 P.M.)

These commuting hours data convincingly show no effect, but are too unstable to help much with estimating the long-term effects. They show a different annual cycle than do the weekend nights or the overall figures, and do not go back far enough to provide an adequate base for estimating this annual cycle with precision.

The use of a highly judgmental category such as "serious injuries" provides an opportunity for pseudo effects owing to a shift in the classifiers' standards. The overall figures are available separately for fatalities, and these show a highly significant effect as strong as that found for the serious injury category or the composite shown in Figure 11A.

More details and the methodological problems are considered in our fuller presentation (Ross, Campbell, & Glass, 1970). One further rule for the use of this design needs emphasizing. The interrupted time series can provide clear evidence of effect only where the reform is introduced with a vigorous abruptness. A gradually introduced reform has little chance of being distinguished from shifts in secular trends or from the cumulative effect of the many other influences impinging during a prolonged period of introduction. In the Breathalyser crackdown, an intense publicity campaign naming the specific starting date preceded the actual crackdown. Although the impact seems primarily due to publicity and fear rather than an actual increase of arrests, an abrupt initiation date was achieved. Had the enforcement effort changed at the moment the Act was passed, with public awareness being built up by subsequent publicity, the resulting data series would have been essentially uninterpretable.

REGRESSION DISCONTINUITY DESIGN

We shift now to social ameliorations that are in short supply, and that therefore cannot be given to all individuals. Such scarcity is inevitable under many circumstances, and can make possible an evaluation of effects that would otherwise be impossible. Consider the heroic Salk poliomyelitis vaccine trials in which some children were given the vaccine while others were given an inert saline placebo injection—and in which many more of these placebo controls would die than would have if they had been given the vaccine. Creation of these placebo controls would have been morally, psychologically, and socially impossible had there been enough vaccine for all. As it was, due to the scarcity, most children that year had to go without the vaccine anyway. The creation of experimental and control groups was the highly moral allocation of that scarcity so as to enable us to learn the true efficacy of the supposed good. The usual medical practice of introducing new cures on a so-called trial basis in general medical practice makes evaluation impossible by confounding prior status with treatment, that is, giving the drug to the most needy or most hopeless. It has the further social bias of giving the supposed benefit to those most assiduous in keeping their medical needs in the attention of the

medical profession, that is, the upper and upper-middle classes. The political stance furthering social experimentation here is the recognition of randomization as the most democratic and moral means of allocating scarce resources (and scarce hazardous duties), plus the moral imperative to further utilize the randomization so that society may indeed learn true value of the supposed boon. This is the ideology that makes possible "true experiments" in a large class of social reforms.

But if randomization is not politically feasible or morally justifiable in a given setting, there is a powerful quasi-experimental design available that allows the scarce good to be given to the most needy or the most deserving. This is the regression discontinuity design. All it requires is strict and orderly attention to the priority dimension. The design originated through an advocacy of a tie-breaking experiment to measure the effects of receiving a fellowship (Thistlethwaite & Campbell, 1960), and it seems easiest to explain it in that light. Consider as in Figure 12, pre-award ability-and-merit dimension, which would have some relation to later success in life

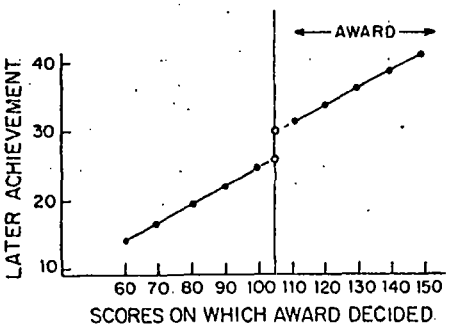


FIG. 12. Tie-breaking experiment and regression discontinuity analysis.

(finishing college, earnings 10 years later, etc.). Those higher on the premeasure are most deserving and receive the award. They do better in later life, but does the award have an effect? It is normally impossible to say because they would have done better in later life anyway. Full randomization of the award was impossible given the stated intention to reward merit and ability. But it might be possible to take a narrow band of ability at the cutting point, to regard all of these persons as tied, and to assign half of them to awards, half to no awards, by means of a tie-breaking randomization.

The tie-breaking rationale is still worth doing, but in considering that design it became obvious that, if the regression of premeasure on later effects were reasonably orderly, one should be able to extrapolate to the results of the tie-breaking experiment by plotting the regression of posttest on pretest separately for those in the award and nonaward regions. If there is no significant difference for these at the decision-point intercept, then the tie-breaking experiment should show no difference. In cases where the tie breakers would show an effect, there should be an abrupt discontinuity in the regression line. Such a discontinuity cannot be explained away by the normal regression of the posttest on pretest, for this normal regression, as extensively sampled within the nonaward area and within the award area, provides no such expectation.

Figure 12 presents, in terms of column means, an instance in which higher pretest scores would have led to higher posttest scores even without the treatment, and in which there is in addition a substantial treatment effect. Figure 13

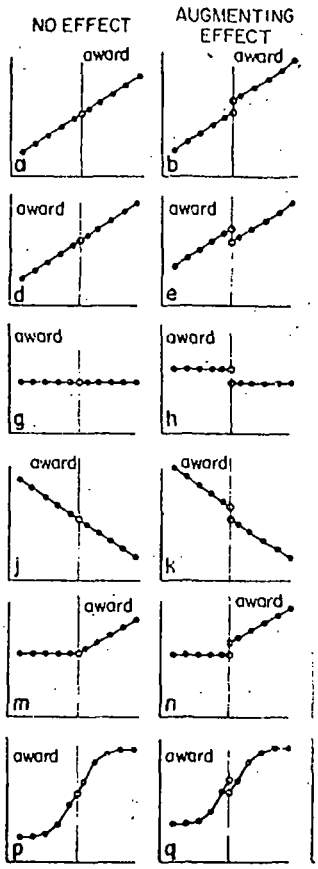


FIG. 13. Illustrative outcom
sion discontinuity analyses.

shows a series of paired outc
on the left to be interpreted
those in the center and on
effect. Note some particular
stances of granting opportu
basis of merit, like 13a and
ure 12), neglect of the bacl
gression of pretest on postt
optimistic pseudo-effects: in
those receiving the award do
later life, though not really
the award. But in social ame
forts, the setting is more ap

· college, earnings 10 years
). Those higher on the pre-
:re most deserving and receive
:. They do better in later life,
the award have an effect? It is
impossible to say because they
ve done better in later life any-
: randomization of the award
ossible given the stated inten-
ward merit and ability. But it
po: ible to take a narrow band
at the cutting point, to regard
se persons as tied, and to assign
them to awards, half to no
y means of a tie-breaking ran-
on.
ie-breaking rationale is still
ing, but in considering that de-
came obvious that, if the regres-
remeasure on later effects were
iy orderly, one should be able to
te to the results of the tie-
· experiment by plotting the re-
of posttest on pretest separately
: in the award and nonaward re-
there is no significant difference
: at the decision-point intercept,
tie-breaking experiment should
: difference. In cases where the
ters would show an effect, there
e an abrupt discontinuity in the
on line. Such a discontinuity can-
xplained away by the normal re-
of the posttest on pretest, for
rmal regression, as extensively
within the nonaward area and
ne award area, provides no such
ion.
: 12 presents, in terms of col-
ans, an instance in which higher
scores would have led to higher
scores even without the treat-
nd in which there is in addition a
tial treatment effect. Figure 13

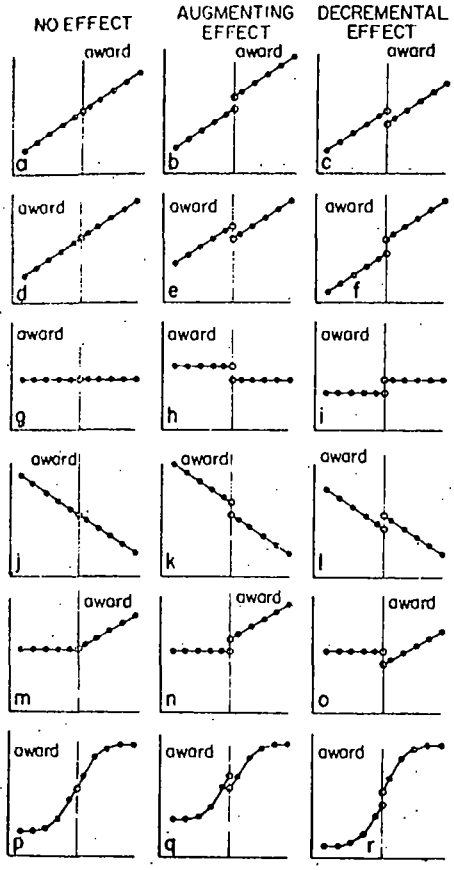


FIG. 13. Illustrative outcomes of regression discontinuity analyses.

shows a series of paired outcomes, those on the left to be interpreted as no effect, those in the center and on the right as effect. Note some particular cases. In instances of granting opportunity on the basis of merit, like 13a and b (and Figure 12), neglect of the background regression of pretest on posttest leads to optimistic pseudo-effects: in Figure 13a, those receiving the award do do better in later life, though not really because of the award. But in social ameliorative efforts, the setting is more apt to be like Figure 13d and e, where neglect of the background regression is apt to make the program look deleterious if no effect, or ineffective if there is a real effect.

The design will of course work just as well or better if the award dimension and the decision base, the pretest measure, are unrelated to the posttest dimension, if it is irrelevant or unfair, as instanced in Figure 13g, h, and i. In such cases the decision base is the functional equivalent of randomization. Negative background relationships are obviously possible, as in Figure 13j, k, and l. In Figure 13, m, n, and o are included to emphasize that it is a jump in intercept at the cutting point that shows effect, and that differences in slope without differences at the cutting point are not acceptable as evidences of effect. This becomes more obvious if we remember that in cases like m, a tie-breaking randomization experiment would have shown no difference. Curvilinear background relationships, as in Figure 13p, q, and r, will provide added obstacles to clear inference in many instances, where sampling error could make Figure 13p look like 13b.

As further illustration, Figure 14 provides computer-simulated data, showing individual observations and fitted regression lines, in a fuller version of the no-effect outcome of Figure 13a. Figure 15 shows an outcome with effect. These have been generated[5] by assigning to each individual a weighted normal random number as a "true score," to which is added a weighted independent "error" to generate the "pretest." The "true score" plus another independent "error" produces the "posttest" in no-effect cases such as Figure 14. In treatment-effect simulations, as in Figure 15, there are

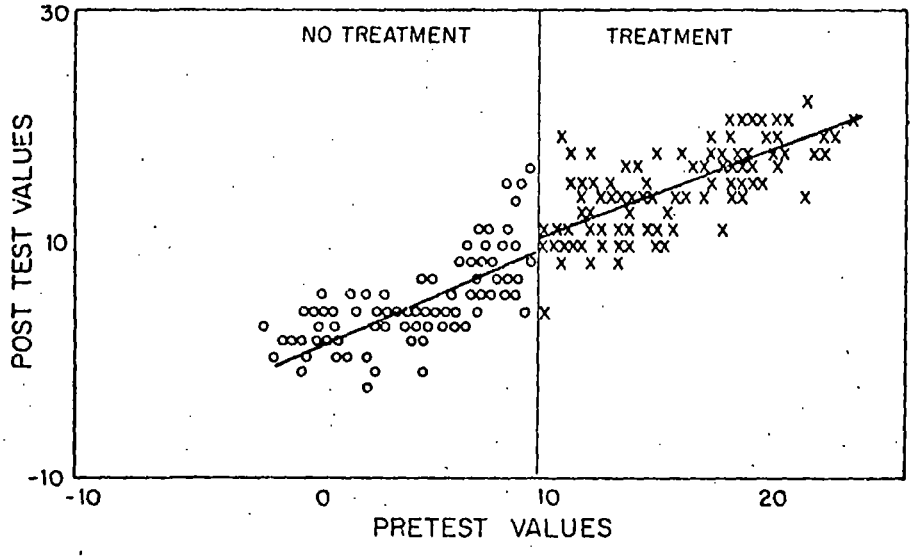


FIG. 14. Regression discontinuity design: No effect.

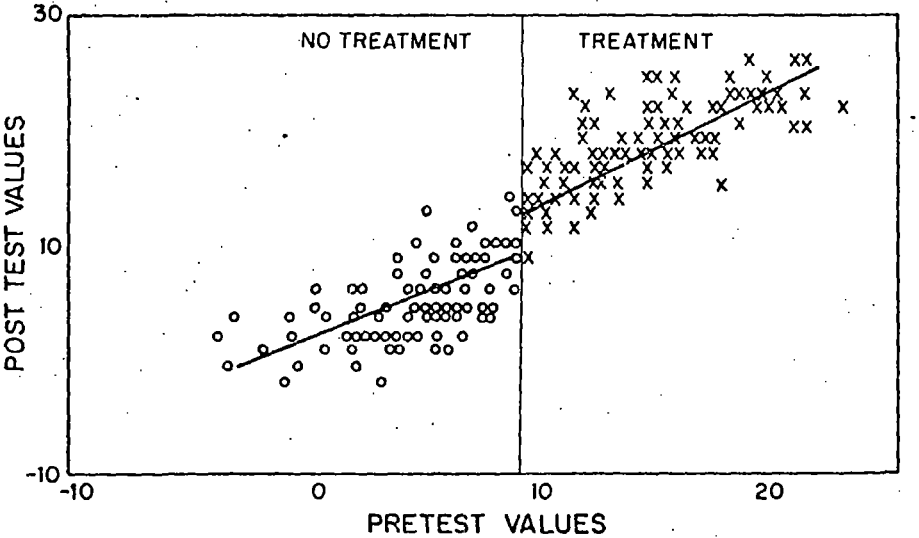


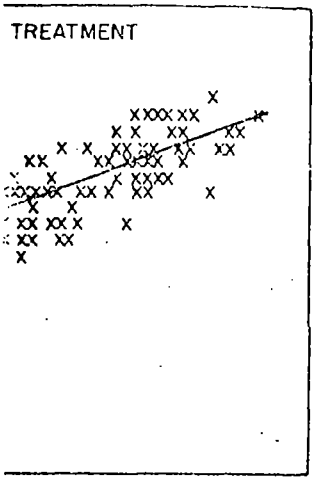
FIG. 15. Regression discontinuity design: Genuine effect.

added into the posttest "effects points" for all "treated" cases, that is, those above the cutting point on the pretest score.

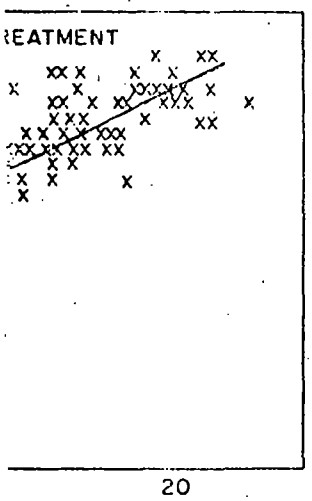
This design could be used in a number of settings. Consider Job Training Corps applicants, in larger number than the program can accommodate, with eligibility determined by need. The setting would be as in Figure 13d and e. The base-line decision dimension could be per capita family income, with those

at below the cutoff getting
outcome dimension could b
of withholding tax withhe
later, or the percentage dr
ployment insurance, these f
ures being provided from
Data Bank in response to ca
cial security numbers fed in
dividual anonymity bein
without any real invasion
by the technique of Mutua
Data Banks. While the pl
could be named, there is n
they be named. In a classic
ment on tax complianc
Schwartz and the Bureau
Revenue have managed to
sets of personally identified
and tax-return data so tha
analyses such as these can be
out the separate custodians c
terview or tax returns learni
responding data for speci
(Schwartz & Orleans, 196
Schwartz & Skolnick, 1963).

Applied to the Job Trainir
lustration, it would work as fc
arate lists of job-corps appli
social security numbers) wo
pared for every class inter
capita family income. To ea
lists an alphabetical designa
be assigned at random. (Thus
per week list might be
$11.00, C, $12.00, Z, $13.00,
N, etc.) These lists would be
ternal Revenue, without th
Revenue personnel being ab
anything interpretable about
eeship status or family in
Internal Revenue statisticians
cate the withholding tax co
each person on each list, but
return the data in that form.
each list, only the withh



3



t.

cants, in larger number than
n can accommodate, with
etermined by need. The set-
be as in Figure 13d and e.
e decision dimension could
a family income, with those

at below the cutoff getting training. The outcome dimension could be the amount of withholding tax withheld two years later, or the percentage drawing unemployment insurance, these follow-up figures being provided from the National Data Bank in response to categorized social security numbers fed in, without individual anonymity being breached, without any real invasion of privacy—by the technique of Mutually Insulated Data Banks. While the plotted points could be named, there is no need that they be named. In a classic field experiment on tax compliance, Richard Schwartz and the Bureau of Internal Revenue have managed to put together sets of personally identified interviews and tax-return data so that statistical analyses such as these can be done, without the separate custodians of either interview or tax returns learning the corresponding data for specific persons (Schwartz & Orleans, 1967; see also Schwartz & Skolnick, 1963).

Applied to the Job Training Corps illustration, it would work as follows: Separate lists of job-corps applicants (with social security numbers) would be prepared for every class interval on per capita family income. To each of these lists an alphabetical designation would be assigned at random. (Thus the $10.00 per week list might be labeled *M*; $11.00, *C*, $12.00, *Z*, $13.00, *Q*, $14.00, *N*, etc.) These lists would be sent to Internal Revenue, without the Internal Revenue personnel being able to learn anything interpretable about their traineeship status or family income. The Internal Revenue statisticians would locate the withholding tax collected for each person on each list, but would not return the data in that form. Instead, for each list, only the withholding tax amounts would be listed, and these in a newly randomized order. These would be returned to Job Corps research, who could use them to plot a graph like Figures 10 or 11, and do the appropriate statistical analyses by retranslating the alphabetical symbols into meaningful base-line values. But, within any list, they would be unable to learn which value belonged to which person. (To insure this effective anonymity, it could be specified that no lists shorter than 100 persons be used, the base-line intervals being expanded if necessary to achieve this.)

Manniche and Hayes (1957) have spelled out how a broker can be used in a two-staged matching of doubly coded data. Kaysen (1967) and Sawyer and Schechter (1968) have wise discussions of the more general problem.

What is required of the administrator of a scarce ameliorative commodity to use this design? Most essential is a sharp cutoff point on a decision-criterion dimension, on which several other qualitatively similar analytic cutoffs can be made both above and below the award cut. Let me explain this better by explaining why National Merit scholarships were unable to use the design for their actual fellowship decision (although it has been used for their Certificate of Merit). In their operation, diverse committees make small numbers of award decisions by considering a group of candidates and then picking from them the *N* best to which to award the *N* fellowships allocated them. This provides one cutting point on an unspecified pooled decision base, but fails to provide analogous potential cutting points above and below. What could be done is for each committee to collectively rank its group of 20 or so candidates. The

top N would then receive the award. Pooling cases across committees, cases could be classified according to number of ranks above and below the cutting point, these other ranks being analogous to the award-nonaward cutting point as far as regression onto posttreatment measures was concerned. Such group ranking would be costly of committee time. An equally good procedure, if committees agreed, would be to have each member, after full discussion and freedom to revise, give each candidate a grade, A+, A, A−, B+, B, etc., and to award the fellowships to the N candidates averaging best on these ratings, with no revisions allowed after the averaging process. These ranking or rating units, even if not comparable from committee to committee in range of talent, in number of persons ranked, or in cutting point, could be pooled without bias as far as a regression discontinuity is concerned, for that range of units above and below the cutting point in which all committees were represented.

It is the dimensionality and sharpness of the decision criterion that is at issue, not its components or validity. The ratings could be based upon nepotism, whimsey, and superstition and still serve. As has been stated, if the decision criterion is utterly invalid we approach the pure randomness of a true experiment. Thus the weakness of subjective committee decisions is not their subjectivity, but the fact that they provide only the one cutting point on their net subjective dimension. Even in the form of average ratings the recommended procedures probably represent some slight increase in committee work load. But this could be justified to the decision committees by the fact that through refusals, etc., it cannot be known at the time of the committee meeting the exact number to whom the fellowship can be offered. Other costs at the planning time are likewise minimal. The primary additional burden is in keeping as good records on the nonawardees as on the awardees. Thus at a low cost, an experimental administrator can lay the groundwork for later scientific follow-ups, the budgets for which need not yet be in sight.

Our present situation is more apt to be one where our pretreatment measures, aptitude measures, reference ratings, etc., can be combined via multiple correlation into an index that correlates highly but not perfectly with the award decision. For this dimension there is a fuzzy cutoff point. Can the design be used in this case? Probably not. Figure 16 shows the pseudo-effect possible if the award decision contributes any valid variance to the quantified pretest evidence, as it usually will. The award regression rides above the nonaward regression just because of that valid variance in this simulated case, there being no true award effect at all. (In simulating this case, the award decision has been based upon a composite of true score plus an independent award error.) Figure 17 shows a fuzzy cutting point plus a genuine award effect.[6] The recommendation to the administrator is clear: aim for a sharp cutting point on a quantified decision criterion. If there are complex rules for eligibility, only one of which is quantified, seek out for follow-up that subset of persons for whom the quantitative dimension was determinate. If political patronage necessitates some decisions inconsistent with a sharp cutoff, record these cases under the heading "qualitative decision rule" and keep them out of your experimental analysis.

Almost all of our ameliorative pro-

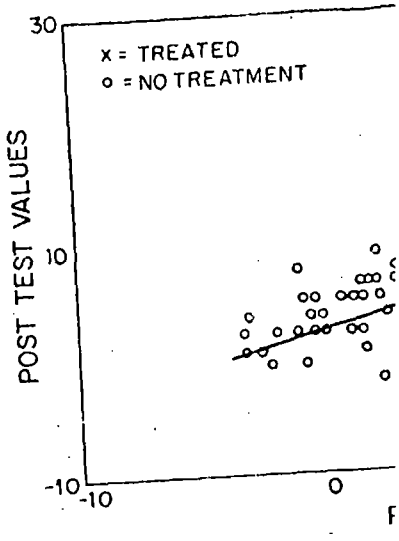


FIG. 16. Regression discontinuity effect only.

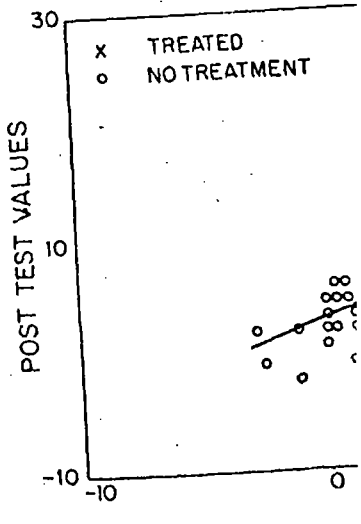


FIG. 17. Regression discontin plus pseudo treatment effects.

ing the exact number to whom
vship can be offered. Other costs
anning time are likewise mini-
e primary additional burden is
ng as good records on the non-
as on the awardees. Thus at a
, an experimental adminis-
o lay the groundwork for later
follow-ups, the budgets for
ed not yet be in sight.
resent situation is more apt to
where our pretreatment meas-
itude measures, reference rat-
can be combined via multiple
n into an index that correlates
t not perfectly with the award
For this dimension there is a
off point. Can the design be
nis case? Probably not. Figure
the pseudo-effect possible if
decision contributes any valid
to the quantified pretest evi-
it usually will. The award
rides above the nonaward re-
ust because of that valid var-
nis simulated case, there being
ward effect at all. (In simu-
s case, the award decision has
d upon a composite of true
an independent award error.)
shows a fuzzy cutting point
mine award effect.[6] The recom-
n to the administrator is clear:
sharp cutting point on a quan-
sion criterion. If there are com-
s for eligibility, only one of
uantified, seek out for follow-
bset of persons for whom the
ve dimension was determinate.
l patronage necessitates some
inconsistent with a sharp cut-
these cases under the heading
e decision rule" and keep
f your experimental analysis.
all of our ameliorative pro-

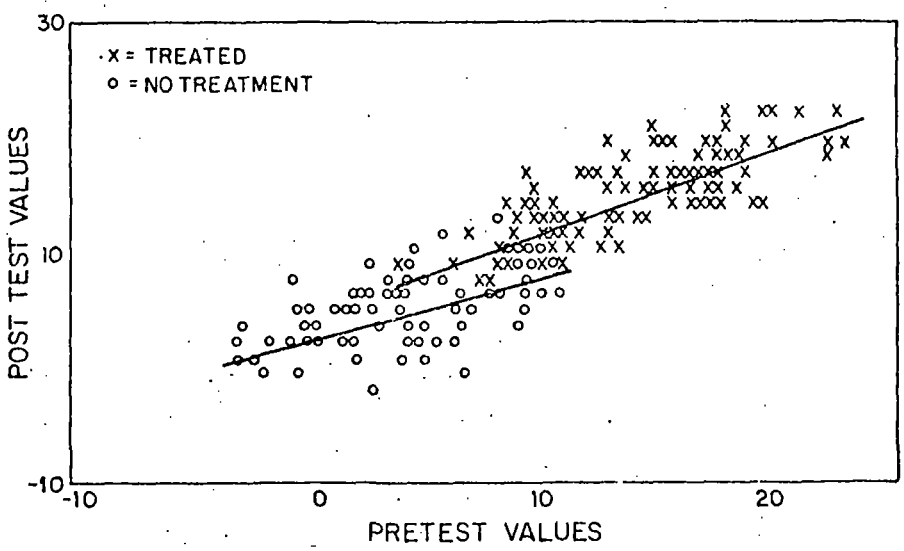


FIG. 16. Regression discontinuity design: Fuzzy cutting point, pseudo treatment effect only.

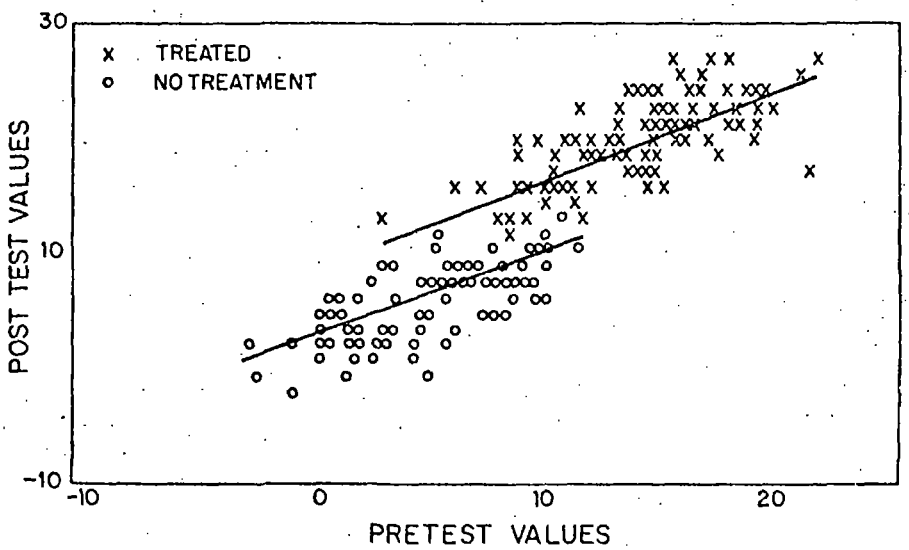


FIG. 17. Regression discontinuity design: Fuzzy cutting point, with real treatment plus pseudo treatment effects.

grams designed for the disadvantaged could be studied via this design, and so too some major governmental actions affecting the lives of citizens in ways we do not think of as experimental. For example, for a considerable period, quantitative test scores have been used to call up for military service or reject as unfit at the lower ability range. If these cutting points, test scores, names, and social security numbers have been recorded for a number of steps both above and below the cutting point, we could make elegant studies of the effect of military service on later withholding taxes, mortality, number of dependents, etc.

This illustration points to one of the threats to external validity of this design, or of the tie-breaking experiment. The effect of the treatment has only been studied for that narrow range of talent near the cutting point, and generalization of the effects of military service, for example, from this low ability level to the careers of the most able would be hazardous in the extreme. But in the draft laws and the requirements of the military services there may be other sharp cutting points on a quantitative criterion that could also be used. For example, those over 6 feet 6 inches are excluded from service. Imagine a five-year-later follow-up of draftees grouped by inch in the 6 feet 1 inch to 6 feet 5 inches range, and a group of their counterparts who would have been drafted except for their heights, 6 feet 6 inches to 6 feet 10 inches. (The fact that the other grounds of deferment might not have been examined by the draft board would be a problem here, but probably not insurmountable.) That we should not expect height in this range to have any relation to later-life variables is not at all a weakness of this design, and if we have indeed a subpopulation for which there is a sharp numerical cutting point, an internally valid measure of effects would result. Deferment under the present system is an unquantified committee decision. But just as the sense of justice of United States soldiers was quantified through paired comparisons of cases into an acceptable Demobilization Points system at the end of World War II (Guttman, 1946; Stouffer, 1949), so a quantified composite index of deferment priority could be achieved and applied as uniform justice across the nation, providing another numerical cutting point.

In addition to the National Data Bank type of indicators, there will be occasions in which new data collections as by interview or questionnaire are needed. For these there is the special problem of uneven cooperation that would be classified as instrumentation error. In our traditional mode of thinking, completeness of description is valued more highly than comparability. Thus if, in a fellowship study, a follow-up mailed out from the fellowship office would bring a higher return from past winners, this might seem desirable even if the nonawardees' rate of response was much lower. From the point of view of quasi-experimentation, however, it would be better to use an independent survey agency and a disguised purpose, achieving equally low response rates from both awardees and nonawardees, and avoiding a regression discontinuity in cooperation rate that might be misinterpreted as a discontinuity in more important effects.

RANDOMIZED CONTROL GROUP EXPERIMENTS

Experiments with randomization tend to be limited to the laboratory and agricultural experiment station. But this certainly need not be so. The r
unit may be persons, famili
or larger administrative uni
tical purposes the random
should be numerous, and l
small. But for reasons of ex
ity, including reactive arran
randomization units should
on the basis of the units of
tive access. Where policies
istered through individual
tacts, randomization at the
may be often inconspicuous
with the clients unaware t
ones of them are getting di
ments. But for most soc
larger administrative units
volved, such as classrooms,
ies, counties, or states. W
develop the political posture
ogies that make randomizat
levels possible.

"Pilot project" is a useful
in our political vocabulary.
a trial program that, if it w
spread to other areas. By m
tual practice in this regard,
ing outside of the popular
ing of the term, a valuable e
ideology could be develope
areas selected for pilot p.oj
public worries about this, it
sumes a lobbying process i
greater needs of some areas
consideration, political pow
pediency being others. W
lating the public tolerance o
could probably devise a syste
the usual lobbying decided
areas eligible for a formal p
that would make final choi
matched pairs. Such decision
as the drawing of lots have
esteemed position since tim
rial (e.g., Aubert, 1959). At
time, record keeping for pi

:here is a sharp numerical cutting
an internally valid measure of ef-
ould result. Deferment under the
system is an unquantified com-
decision. But just as the sense of
of United States soldiers was
:d through paired comparisons
into an acceptable Demobiliza-
:nts system at the end of World
(Guttman, 1946; Stouffer,
so a quantified composite index
:ment priority could be achieved
:ied as uniform justice across the
providing another numerical cut-
:at.

:dition to the National Data
:e of indicators, there will be oc-
in which new data collections
nterview or questionnaire are
For these there is the special
of uneven cooperation that
:e classified as instrumentation
our traditional mode of think-
:pleteness of description is val-
:e highly than comparability.
in a fellowship study, a follow-
:d out from the fellowship office
ring a higher return from past
this might seem desirable even
:nawardees' rate of response was
:wer. From the point of view
:-experimentation, however, it
: better to use an independent
gency and a disguised purpose,
equally low response rates
:h awardees and nonawardees.
:ding a regression discontinuity
:ation rate that might be mis-
:d as a discontinuity in more
t effects.

:IZED CONTROL GROUP :ENTS

:ments with randomization tend
:ted to the laboratory and agri-
:xperiment station. But this cer-
tainly need not be so. The randomization unit may be persons, families, precincts, or larger administrative units. For statistical purposes the randomization units should be numerous, and hence ideally small. But for reasons of external validity, including reactive arrangements, the randomization units should be selected on the basis of the units of administrative access. Where policies are administered through individual client contacts, randomization at the person level may be often inconspicuously achieved, with the clients unaware that different ones of them are getting different treatments. But for most social reforms, larger administrative units will be involved, such as classrooms, schools, cities, counties, or states. We need to develop the political postures and ideologies that make randomization at these levels possible.

"Pilot project" is a useful term already in our political vocabulary. It designates a trial program that, if it works, will be spread to other areas. By modifying actual practice in this regard, without going outside of the popular understanding of the term, a valuable experimental ideology could be developed. How are areas selected for pilot projects? If the public worries about this, it probably assumes a lobbying process in which the greater needs of some areas are only one consideration, political power and expediency being others. Without violating the public tolerance or intent, one could probably devise a system in which the usual lobbying decided upon the areas eligible for a formal public lottery that would make final choices between matched pairs. Such decision procedures as the drawing of lots have had a justly esteemed position since time immemorial (e.g., Aubert, 1959). At the present time, record keeping for pilot projects tends to be limited to the experimental group only. In the experimental ideology, comparable data would be collected on designated controls. (There are of course exceptions, as in the heroic Public Health Service fluoridation experiments, in which the teeth of Oak Park children were examined year after year as controls for the Evanston experimentals [Blayney & Hill, 1967].)

Another general political stance making possible experimental social amelioration is that of *staged innovation.* Even though by intent a new reform is to be put into effect in all units, the logistics of the situation usually dictate that simultaneous introduction is not possible. What results is a haphazard sequence of convenience. Under the program of staged innovation, the introduction of the program would be deliberately spread out, and those units selected to be first and last would be randomly assigned (perhaps randomization from matched pairs), so that during the transition period the first recipients could be analyzed as experimental units, the last recipients as controls. A third ideology making possible true experiments has already been discussed: randomization as the democratic means of allocating scarce resources.

This article will not give true experimentation equal space with quasi-experimentation only because excellent discussions of, and statistical consultation on, true experimentation are readily available. True experiments should almost always be preferred to quasi-experiments where both are available. Only occasionally are the threats to external validity so much greater for the true experiment that one would prefer a quasi-experiment. The uneven allocation of space here should not be read as indicating otherwise.

MORE ADVICE FOR TRAPPED ADMINISTRATORS

But the competition is not really between the fairly interpretable quasi-experiments here reviewed and "true" experiments. Both stand together as rare excellencies in contrast with a morass of obfuscation and self-deception. Both to emphasize this contrast, and again as guidelines for the benefit of those trapped administrators whose political predicament will not allow the risk of failure, some of these alternatives should be mentioned.

Grateful testimonials. Human courtesy and gratitude being what it is, the most dependable means of assuring a favorable evaluation is to use voluntary testimonials from those who have had the treatment. If the spontaneously produced testimonials are in short supply, these should be solicited from the recipients with whom the program is still in contact. The rosy glow resulting is analogous to the professor's impression of his teaching success when it is based solely upon the comments of those students who come up and talk with him after class. In many programs, as in psychotherapy, the recipient, as well as the agency, has devoted much time and effort to the program and it is dissonance reducing for himself, as well as common courtesy to his therapist, to report improvement. These grateful testimonials can come in the language of letters and conversation, or be framed as answers to multiple-item "tests" in which a recurrent theme of "I am sick," "I am well," "I am happy," "I am sad" recurs. Probably the testimonials will be more favorable as: (*a*) the more the evaluative meaning of the response measure is clear to the recipient—it is completely clear in most personality, adjustment, morale, and attitude tests; (*b*) the more directly the recipient is identified by name with his answer; (*c*) the more the recipient gives the answer directly to the therapist or agent of reform; (*d*) the more the agent will continue to be influential in the recipient's life in the future; (*e*) the more the answers deal with feelings and evaluations rather than with verifiable facts; and (*f*) the more the recipients participating in the evaluation are a small and self-selected or agent-selected subset of all recipients. Properly designed the grateful testimonial method can involve pretests as well as posttests, and randomized control groups as well as experimentals, for there are usually no placebo treatments, and the recipients know when they have had the boon.

Confounding selection and treatment. Another dependable tactic bound to give favorable outcomes is to confound selection and treatment, so that in the published comparison those receiving the treatment are also the more able and well placed. The often-cited evidence of the dollar value of a college education is of this nature—all careful studies show that most of the effect, and of the superior effect of superior colleges, is explainable in terms of superior talents and family connections, rather than in terms of what is learned or even the prestige of the degree. Matching techniques and statistical partialings generally undermatch and do not fully control for the selection differences—they introduce regression artifacts confusable as treatment effects.

There are two types of situations that must be distinguished. First, there are those treatments that are given to the most promising, treatments like a college education which are regularly given to those who need it least. For these, the later concomitants of the g
lection operate in the same
the treatment: those mc
achieve anyway get into the
likely to produce later achi
these settings, the trapped
should use the pooled mea
treated, comparing it with
all untreated, although in tl
most any comparison an
might hit upon would be
favor.

At the other end of the
tinuum are those remedia
given to those who need it
the later concomitants of th
selection are poorer success
Training Corps example,
parisons of the later unemp
of those who received the t
those who did not are in g
against showing an advan
training. Here the trapped a
must be careful to seek ou
special comparisons biasing
his favor. For training prog
Operation Head Start and t
grams, a useful solution is to
later success of those who cc
training program with thos
invited but never showed
who came a few times a
our. By regarding only thos
plete the program as "trainec
the others as controls, one is
conscientiousness, stable a
ing family backgrounds, er
the training activity, ability,
tion to get ahead in the wo
tors promising well for fur
ment even if the remedial
valueless. To apply this tacti
in the Job Training Corps,
have to eliminate from the sc
trol group all those who qu

:de tests; (*b*) the more directly
ient is identified by name with
:r; (*c*) the more the recipient
answer directly to the therapist
of reform; (*d*) the more the
il continue to be influential in
:nt's life in the future; (*e*) the
answers deal with feelings and
as rather than with verifiable
! (*f*) the more the recipients
:ing in the evaluation are a
: self-selected or agent-selected
:il recipients. Properly designed
:ul testimonial method can in-
:tests as well as posttests, and
ed control groups as well as
ntals, for there are usually no
:reatments, and the recipients
en they have had the boon.

:nding selection and treatment.
iependable tactic bound to give
outcomes is to confound selec-
treatment, so that in the pub-
mparison those receiving the
are also the more able and
:d. The often-cited evidence of
· value of a college education
:s nature—all careful studies
most of the effect, and of the
fiect of superior colleges, is ex-
in terms of superior talents
ty connections, rather than in
what is learned or even the
:f the degree. Matching tech-
d statistical partialings gener-
match and do not fully control
:ection differences—they intro-
:ession artifacts confusable as
effects.

:re two types of situations that
distinguished. First, there are
uments that are given to the
nising, treatments like a col-
tion which are regularly given
ho need it least. For these, the later concomitants of the grounds of selection operate in the same direction as the treatment: those most likely to achieve anyway get into the college most likely to produce later achievement. For these settings, the trapped administrator should use the pooled mean of all those treated, comparing it with the mean of all untreated, although in this setting almost any comparison an administrator might hit upon would be biased in his favor.

At the other end of the talent continuum are those remedial treatments given to those who need it most. Here the later concomitants of the grounds of selection are poorer success. In the Job Training Corps example, casual comparisons of the later unemployment rate of those who received the training with those who did not are in general biased against showing an advantage to the training. Here the trapped administrator must be careful to seek out those few special comparisons biasing selection in his favor. For training programs such as Operation Head Start and tutoring programs, a useful solution is to compare the later success of those who completed the training program with those who were invited but never showed plus those who came a few times and dropped out. By regarding only those who complete the program as "trained" and using the others as controls, one is selecting for conscientiousness, stable and supporting family backgrounds, enjoyment of the training activity, ability, determination to get ahead in the world—all factors promising well for future achievement even if the remedial program is valueless. To apply this tactic effectively in the Job Training Corps, one might have to eliminate from the so-called control group all those who quit the training program because they had found a job—but this would seem a reasonable practice and would not blemish the reception of a glowing progress report.

These are but two more samples of well-tried modes of analysis for the trapped administrator who cannot afford an honest evaluation of the social reform he directs. They remind us again that we must help create a political climate that demands more rigorous and less self-deceptive reality testing. We must provide political stances that permit true experiments, or good quasi-experiments. Of the several suggestions toward this end that are contained in this article, the most important is probably the initial theme: Administrators and parties must advocate the importance of the problem rather than the importance of the answer. They must advocate experimental sequences of reforms, rather than one certain cure-all, advocating Reform A with Alternative B available to try next should an honest evaluation of A prove it worthless or harmful.

MULTIPLE REPLICATION IN ENACTMENT

Too many social scientists expect single experiments to settle issues once and for all. This may be a mistaken generalization from the history of great crucial experiments in physics and chemistry. In actuality the significant experiments in the physical sciences are replicated thousands of times, not only in deliberate replication efforts, but also as inevitable incidentals in successive experimentation and in utilizations of those many measurement devices (such as the galvanometer) that in their own operation embody the principles of classic experiments. Because we social scien-

tists have less ability to achieve "experimental isolation," because we have good reason to expect our treatment effects to interact significantly with a wide variety of social factors many of which we have not yet mapped, we have much greater needs for replication experiments than do the physical sciences.

The implications are clear. We should not only do hard-headed reality testing in the initial pilot testing and choosing of which reform to make general law; but once it has been decided that the reform is to be adopted as standard practice in all administrative units, we should experimentally evaluate it in each of its implementations (Campbell, 1967).

CONCLUSIONS

Trapped administrators have so committed themselves in advance to the efficacy of the reform that they cannot afford honest evaluation. For them, favorably biased analyses are recommended, including capitalizing on regression, grateful testimonials, and confounding selection and treatment. *Experimental administrators* have justified the reform on the basis of the importance of the problem, not the certainty of their answer, and are committed to going on to other potential solutions if the one first tried fails. They are therefore not threatened by a hard-headed analysis of the reform. For such, proper administrative decisions can lay the base for useful experimental or quasi-experimental analyses. Through the ideology of allocating scarce resources by lottery, through the use of staged innovation, and through the pilot project, true experiments with randomly assigned control groups can be achieved. If the reform must be introduced across the board, the interrupted time-series design is available. If there are similar units under independent administration, a control series design adds strength. If a scarce boon must be given to the most needy or to the most deserving, quantifying this need or merit makes possible the regression discontinuity analysis.

NOTES

[1] This list has been expanded from the major previous presentations by the addition of *Instability* (but see Campbell, 1968; Campbell & Ross, 1968). This has been done in reaction to the sociological discussion of the use of tests of significance in nonexperimental or quasi-experimental research (e.g., Selvin, 1957; and as reviewed by Galtung, 1967, pp. 358–389). On the one hand, I join with the critics in criticizing the exaggerated status of "statistically significant differences" in establishing convictions of validity. Statistical tests are relevant to at best 1 out of 15 or so threats to validity. On the other hand, I join with those who defend their use in situations where randomization has not been employed. Even in those situations, it is relevant to say or to deny, "This is a trivial difference. It is of the order that would have occurred frequently *had* these measures been assigned to these classes solely by chance." Tests of significance, making use of random reassignments of the actual scores, are particularly useful in communicating this point.

[2] This list has been lengthened from previous presentations to make more salient Threats 5 and 6 which are particularly relevant to social experimentation. Discussion in previous presentations (Campbell, 1957, pp. 309–310; Campbell & Stanley, 1963, pp. 203–204) had covered these points, but they had not been included in the checklist.

[3] No doubt the public and press shared the Governor's special alarm over the 1955 death toll. This differentia
which the dampening eff
the prior trend. Insofar a
ponent to the autocorrela
probably be regarded as a
sion effect is less as the p
degree insofar as this cor
time series would represe
to negative correlation ac
correlation would be repre
to the other.)

[4] Wilson's inconsistency
records are ably document
York City in 1965 a crime
licized improvement in rec

[5] J. Sween & D. T. Cam
and fuzzy regression-discor

[6] There are some subtle
one had enough cases. Th
columns for a true effects
been a sharp cutting point
no discontinuity in the no-
albeit an underestimated d
point and treated ones bel
ening should be estimable
are hopes for the future.

REFERENCES

Aubert, V. Chance in social
Bauer, R. M. *Social indicat*
Blayney, J. R., & Hill, I.
Dental Association (Spe
Box, G. E. P., & Tiao, G. C
1965, 52, 181–192.
Campbell, D. T. Factors r
chological Bulletin, 195
Campbell, D. T. From des
periments. In C. W. Ha
sity of Wisconsin Press,
Campbell, D. T. Administ
measures. In J. C. Stanle
sis. Chicago: Rand McN
Campbell, D. T. Quasi-ex
clopedia of the Social Sc
259–263.
Campbell, D. T., & Fiske,
trait-multimethod matri
Campbell, D. T., & Ross,
data in quasi-experiment

grateful testimonials, and
ing selection and treatment.
ntal administrators have justi-
reform on the basis of the
ce of the problem, not the cer-
their answer, and are com-
going on to other potential
if the one first tried fails. They
fore not threatened by a hard-
nalysis of the reform. For such,
dministrative decisions can lay
for useful experimental or
erimental analyses. Through
logy of allocating scarce re-
by lottery, through the use of
novation, and through the pilot
rue experiments with randomly
control groups can be achieved.
form must be introduced across
d, the interrupted time-series
available. If there are similar
ier independent administration,
series design adds strength. If
boon must be given to the most
to the most deserving, quanti-
's need or merit makes possible
ssion discontinuity analysis.

ous presentations by the addition
Goss, 1968). This has been done
tests of significance in nonexperi-
57; and as reviewed by Galtung,
critics in criticizing the exagger-
tablishing convictions of validity.
threats to validity. On the other
ons where randomization has not
o say or to deny, "This is a trivial
requently *had* these measures been
nificance, making use of random
in communicating this point.
sentations to make more salient
al experimentation. Discussion in
Campbell & Stanley, 1963, pp.
een included in the checklist.
or's special alarm over the 1955

death toll. This differential reaction could be seen as a negative feedback servosystem in which the dampening effect was proportional to the degree of upward deviation from the prior trend. Insofar as such alarm reduces traffic fatalities, it adds a negative component to the autocorrelation, increasing the regression effect. This component should probably be regarded as a rival cause or treatment rather than as artifact. (The regression effect is less as the positive autocorrelation is higher, and will be present to some degree insofar as this correlation is less than positive unity. Negative correlation in a time series would represent regression beyond the mean, in a way not quite analogous to negative correlation across persons. For an autocorrelation of Lag 1, high negative correlation would be represented by a series that oscillated maximally from one extreme to the other.)

[4] Wilson's inconsistency in utilization of records and the political problem of relevant records are ably documented in Kamisar (1964). Etzioni (1968) reports that in New York City in 1965 a crime wave was proclaimed that turned out to be due to an unpublicized improvement in record keeping.

[5] J. Sween & D. T. Campbell, Computer programs for simulating and analyzing sharp and fuzzy regression-discontinuity experiments. In preparation.

[6] There are some subtle statistical clues that might distinguish these two instances if one had enough cases. There should be increased pooled column variance in the mixed columns for a true effects case. If the data are arbitrarily treated as though there had been a sharp cutting point located in the middle of the overlap area, then there should be no discontinuity in the no-effect case, and some discontinuity in the case of a real effect, albeit an underestimated discontinuity, since there are untreated cases above the cutting point and treated ones below, dampening the apparent effect. The degree of such dampening should be estimable, and correctable, perhaps by iterative procedures. But these are hopes for the future.

REFERENCES

Aubert, V. Chance in social affairs. *Inquiry*, 1959, 2, 1–24.

Bauer, R. M. *Social indicators*. Cambridge, Mass.: M.I.T. Press, 1966.

Blayney, J. R., & Hill, I. N. Fluorine and dental caries. *The Journal of the American Dental Association* (Special Issue), 1967, 74, 233–302.

Box, G. E. P., & Tiao, G. C. A change in level of a nonstationary time series. *Biometrika*, 1965, 52, 181–192.

Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 1957, 54, 297–312.

Campbell, D. T. From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.

Campbell, D. T. Administrative experimentation, institutional records, and nonreactive measures. In J. C. Stanley (Ed.), *Improving experimental design and statistical analysis*. Chicago: Rand McNally, 1967.

Campbell, D. T. Quasi-experimental design. In D. L. Sills (Ed.), *International Encyclopedia of the Social Sciences*. New York: Macmillan and Free Press, 1968, Vol. 5, 259–263.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81–105.

Campbell, D. T., & Ross, H. L. The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. *Law and Society Review*, 1968, 3(1), 33–53.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963. (Reprinted as *Experimental and quasi-experimental design for research.* Chicago: Rand McNally, 1966.)

Chapin, F. S. *Experimental design in sociological research.* New York: Harper, 1947.

Etzioni, A. "Shortcuts" to social change? *The Public Interest,* 1968, 12, 40–51.

Etzioni, A., & Lehman, E. W. Some dangers in "valid" social measurement. *Annals of the American Academy of Political and Social Science,* 1967, 373, 1–15.

Galtung, J. *Theory and methods of social research.* Oslo: Universitetsforloget; London: Allen & Unwin; New York: Columbia University Press, 1967.

Glass, G. V. Analysis of data on the Connecticut speeding crackdown as a time-series quasi-experiment. *Law and Society Review,* 1968, 3(1), 55–76.

Glass, G. V., Tiao, G. C., & Maguire, T. O. Analysis of data on the 1900 revision of the German divorce laws as a quasi-experiment. *Law and Society Review,* 1969, in press.

Greenwood, E. *Experimental sociology: A study in method.* New York: King's Crown Press, 1945.

Gross, B. M. *The state of the nation: Social system accounting.* London: Tavistock Publications, 1966. (Also in R. M. Bauer, *Social indicators.* Cambridge, Mass.: M.I.T. Press, 1966.)

Gross, B. M. (Ed.) Social goals and indicators. *Annals of the American Academy of Political and Social Science,* 1967. 371, Part 1, May, pp. i–iii and 1–177; Part 2, September, pp. i–iii and 1–218.

Guttman, L. An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics,* 1946, 17, 144–163.

Hyman, H. H., & Wright, C. R. Evaluating social action programs. In P. F. Lazarsfeld, W. H. Sewell, & H. L. Wilensky (Eds.), *The uses of sociology.* New York: Basic Books, 1967.

Kamisar, Y. The tactics of police-persecution oriented critics of the courts. *Cornell Law Quarterly,* 1964, 49, 458–471.

Kaysen, C. Data banks and dossiers. *The Public Interest,* 1967, 7, 52–60.

Manniche, E., & Hayes, D. P. Respondent anonymity and data matching. *Public Opinion Quarterly,* 1957, 21(3), 384–388.

Polanyi, M. A society of explorers. In, *The tacit dimension.* (Ch. 3) New York: Doubleday, 1966.

Polanyi, M. The growth of science in society. *Minerva,* 1967, 5, 533–545.

Popper, K. R. *Conjectures and refutations.* London: Routledge and Kegan Paul; New York: Basic Books, 1963.

Rheinstein, M. Divorce and the law in Germany: A review. *American Journal of Sociology,* 1959, 65, 489–498.

Rose, A. M. Needed research on the mediation of labor disputes. *Personnel Psychology,* 1952, 5, 187–200.

Ross, H. L., & Campbell, D. T. The Connecticut speed crackdown: A study of the effects of legal change. In H. L. Ross (Ed.), *Perspectives on the social order: Readings in sociology.* New York: McGraw-Hill, 1968.

Ross, H. L., Campbell, D. T., and Glass, G. V. Determining the social effects of a legal reform: The British "Breathalyser" crackdown of 1967. *American Behavioral Scientist,* 1970, 13, 493–509.

Sawyer, J., & Schechter, H. Computers, privacy, and the National Data Center: The responsibility of social scientists. *American Psychologist,* 1968, 23, 810–818.

Schanck, R. L., & Good
issue. *Public Opinion*
Schwartz, R. D. Field e
tion, 1961, 13, 401–
Schwartz, R. D., & Orle
1967, 34, 274–300.
Schwartz, R. D., & Skolr
In L. Arons & M. Ma
ton-Century-Crofts, 1
Selvin, H. A critique of
Review, 1957, 22, 51
Simon, J. L. The price el
tion. *Econometrica,* 19
Solomon, R. W. An exte
137–150.
Stieber, J. W. *Ten years*
Relations Center, Univ
Stouffer, S. A. The point
The American soldier.
sity Press, 1949.
Suchman, E. A. *Evaluativ*
action programs. New
Sween, J., & Campbell, D.
tests of significance for
able from author, 1965.
Thistlethwaite, D. L., & C
tive to the ex post fact
309–317.
Walker, H. M., & Lev, J. S
Webb, E. J., Campbell, D.
Nonreactive research in
Wolf, E., Lüke, G., & Ha
scheidung in Deutschlan

xperimental designs for re-
research on teaching. Chi-
and quasi-experimental de-

lew York: Harper, 1947.
1968, 12, 40–51.
rial measurement. *Annals of*
57, 373, 1–15.
niversitetsforloget; London:
1967.
z c ckdown as a time-series
55–76.
a on the 1900 revision of the
iety Review, 1969, in press.
l. New York: King's Crown

ing. London: Tavistock Pub-
. Cambridge, Mass.: M.I.T.

of the American Academy of
i–iii and 1–177; Part 2, Sep-

ons and rank order. *Annals of*

programs. In P. F. Lazarsfeld,
sociology. New York: Basic

ics of the courts. *Cornell Law*

967, 7, 52–60.
data matching. *Public Opinion*

n. (Ch. 3) New York: Dou-

57, 5, 533–545.
tledge and Kegan Paul; New

w. *American Journal of Sociol-*

lisputes. *Personnel Psychology,*

crackdown: A study of the ef-
s on the social order: Readings

ning the social effects of a legal
7. *American Behavioral Scien-*

National Data Center: The re-
1968, 23, 810–818.

Schanck, R. L., & Goodman, C. Reactions to propaganda on both sides of a controversial issue. *Public Opinion Quarterly*, 1939, 3, 107–112.

Schwartz, R. D. Field experimentation in sociological research. *Journal of Legal Education*, 1961, 13, 401–410.

Schwartz, R. D., & Orleans, S. On legal sanctions. *University of Chicago Law Review*, 1967, 34, 274–300.

Schwartz, R. D., & Skolnick, J. H. Televised communication and income tax compliance. In L. Arons & M. May (Eds.), *Television and human behavior*. New York: Appleton-Century-Crofts, 1963.

Selvin, H. A critique of tests of significance in survey research. *American Sociological Review*, 1957, 22, 519–527.

Simon, J. L. The price elasticity of liquor in the U.S. and a simple method of determination. *Econometrica*, 1966, 34, 193–205.

Solomon, R. W. An extension of control group design. *Psychological Bulletin*, 1949, 46, 137–150.

Stieber, J. W. *Ten years of the Minnesota Labor Relations Act*. Minneapolis: Industrial Relations Center, University of Minnesota, 1949.

Stouffer, S. A. The point system for redeployment and discharge. In S. A. Stouffer et al., *The American soldier. Vol. 2, Combat and its aftermath*. Princeton: Princeton University Press, 1949.

Suchman, E. A. *Evaluative research: Principles and practice in public service and social action programs*. New York: Russell Sage, 1967.

Sween, J., & Campbell, D. T. A study of the effect of proximally auto-correlated error on tests of significance for the interrupted time-series quasi-experimental design. Available from author, 1965. (Multilith)

Thistlethwaite, D. L., & Campbell, D. T. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 1960, 51, 309–317.

Walker, H. M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. B. *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally, 1966.

Wolf, E., Lüke, G., & Hax, H. *Scheidung und Scheidungsrecht: Grundfragen der Ehescheidung in Deutschland*. Tübigen: J. C. B. Mohr, 1959.

Soc. Sci. & Med. 1972, Vol. 6 pp. 469–490. Pergamon Press. Printed in Great Britain

ON THE SCOPE AND METHODOLOGY OF RESEARCH IN PUBLIC HEALTH PRACTICE

A. F. WESSEN*

Chief, Behavioural Science Unit, World Health Organization, Geneva

Abstract—The purpose of this paper is to present an approach to the problems of public health practice research which will be at once panoramic and specific enough to provide some focus upon issues and methods. Like every paper, this one has biases which reflect the background of its author; it should be stated that this paper reflects the views of a sociologist concerned throughout his professional career with the study of medical care and heavily influenced by the developing interdisciplinary viewpoint evolving in the Division of Research in Epidemiology and Communications Science in the World Health Organization. The paper is comprised of four principal sections. Beginning with a discussion of the kinds of substantive problems and research skills involved in public health practice research, the discussion then proceeds to an analysis of some conceptual problems that such research must face and an enumeration of some of the types of methods which may be of use in solving these. The paper concludes with an assessment of some of the key problems which must be mastered if the potential of this field is to be realized throughout the world.

THE SCOPE OF PUBLIC HEALTH PRACTICE RESEARCH

ONE MUST begin with a brief discussion of what public health practice involves. Most generally, it would seem to denote the process of the application of scientific knowledge of medicine toward the amelioration of the health of the community [1].

Several observations should be made about this definition. *First*, it specifically excludes from public health practice research the problems of enhancing the scientific knowledge of medicine itself. Questions about the aetiology, pathology, and treatment of disease itself are therefore excluded as well as questions concerning the biological characteristics and functioning of the human organism. *Second*, the focus is upon the *process* of the *application* of this knowledge to a population. What public health practitioners need to know is *what use* to make of the scientific knowledge of medicine, *how to organize* the facilities and resources necessary to make this use take place, *how the process* of action and decision involved in application *takes place*, and *what its consequences are*. *Third*, the target of public health practice is always a population, a "community" of defined scope. Intrinsic to the idea of public health practice, therefore, is knowledge about the characteristics and needs of that community. *Fourth*, there is always a *purposive*, even a *normative* character to public health practice. It has a goal: the improvement of levels of health in the community. Therefore, because of this goal (however difficult it may be to specify concretely), it is always possible to evaluate public health practice in terms of criteria of *effectiveness* and *efficiency* [2]. And research in public health practice is often, therefore, directed toward how to *optimize* the degree to which scientific knowledge is really applied and how to minimize the costs entailed in so doing [3]. Finally, public health practice historically has emphasized the provision of

* Now, Professor of Sociology, Brown University, Providence, R.I. The author alone is responsible for views expressed in this article, which was prepared for presentation at a WHO Consultation of Experts in Research in Public Health Practice, Geneva, December 1968.

"organized community effort" [4]. It would thus be distinguished both from the process of providing personal medical care to individuals and from the "unorganized" health activities found within a population—its folk medicine or its consensual standards of hygiene, for example [5].

With all of this delimitation, however, there still remains a very wide range of phenomena to be considered. Public health practice concerns itself with the protection of the community against diseases both through sanitation and other environmental controls and through the promotion of preventive measures among a population. It also is concerned with the provision of services for care of the sick, with their organization and with the procurement and allocation of necessary health resources, both human and material. It is probably true that the environmental control aspects of public health practice, depending heavily as they do upon engineering skills, have tended to constitute a separate sub-discipline within the field. The operational and technical problems entailed in providing personal care services (both preventive and curative) naturally have drawn upon other skills and have become another nucleus of public health practice; this area includes the burgeoning field of "medical care".

It should be clear that public health practice, to be effective, must be comprehensive in scope. It must necessarily draw upon the skills of several disciplines because it deals with many kinds of activities. Thus the whole catalogue of health professions—medicine, epidemiology, dentistry, nursing, health education, sanitary engineering, microbiology and other laboratory professions, veterinary medicine, the mental health professions, nutrition, biostatistics—and a variety of administrative roles, are regularly included in departments of public health, and instruction in most of these areas is given in schools for public health. Except in terms of a common ideology of health promotion, however, it is not always easy to see how these separate disciplines are integrated. The tendency toward fractionation is rife. There are both the basic cleavage between environmental and personal care aspects already mentioned, and the stubborn tendency to separate preventive and curative services [6]. Additionally, there is the pervasive tendency of medicine both to specialize into sub-disciplines and within these to concentrate upon disease-specific programmes. The result is that public health organizations tend to be composed of a set of more or less autonomous categorical programmes or departments which often affect the community in a set of specific but only vaguely related ways. A fundamental issue of public health practice, therefore, is whether and to what degree coordination and integration of activities should be achieved. How can a rational system of planning and administration of health services be attained? This issue constitutes an important focus of contemporary research in public health practice.

Comprehensive as it is, however, public health practice as an applied discipline cannot attain the relative autonomy and systematization of subject matter which is characteristic of the basic or "pure" sciences. Because it must apply biological and physical science knowledge about disease and its control within the context of a community—a social setting—public health practice necessarily is an interface between the biological and physical sciences on the one hand and the social sciences on the other. Since it involves organized action to achieve goals in the community, public health practice must draw upon the knowledge and lore of the fields of administration and management, accounting, and education. As an academic field, therefore, public health practice represents an attempt at a practical synthesis of a number of disparate fields and is constantly dependent upon inputs of new knowledge from all these areas.

As an area of action also, public health practice represents a sub-system which is itself incomplete and dependent on a larger context. Organized public health agencies are most commonly parts of a larger governmental administrative apparatus and partake of the peculiarities and constraints characteristic of this larger system. The governmental apparatus itself is but a component part of the web of groups and organizations that constitute the social structure of a community (or society). These structures operate within an ecological context, constantly adapting to environmental shifts whether these be physical, biological, economic, or cultural in nature. The result of this is that the practitioner of public health practice must constantly adjust his activities to the intelligence available to him concerning the larger social and ecological environments of which his little system is a part. To a much greater degree than the more closed system of the hospital ward, in which a man–machine complex can intervene in the biological problems of patients who are temporarily in an artificially, highly controlled environment, the public health practitioner can never escape the exigencies of having to deal with fast-changing environments which he cannot control. Both at the level of action and thought, therefore, the field of public health practice is a domain which is characterized by heavy dependence on many input sources beyond its control. Its integrity as a field depends upon its ability to systematize and to integrate these inputs.

These characteristics of public health practice as a field go far toward determining the specific character and problems of its research. As opposed to the more routine data gathering activities appropriate to the intelligence function in good administration, public health practice research must be oriented toward producing either new knowledge concerning the heterogeneous set of variables affecting its field or about new methods of practice. Insofar as the field is an applied discipline, the test of the relevance of its research will be its contribution toward enhancing goal achievement in public health practice.

It would seem that the substantive problems faced by public health practice research can be grouped under seven headings. I have attempted to organize these headings in a manner which roughly follows the flow of information which the public health administrator might require if he were designing a health care system. Although theoretically these areas of research might be applicable to problems of environmental health, I have restricted my discussion to the personal care services area; in so doing, not only can the analysis be more sharply focused, but also I follow common practice in the area.

1. *The assessment of levels of health in populations*

In any applied field, the first logical question would appear to be "what, specifically is needed? What is the nature of the gap between current reality and the goals one can reasonably put forward?" For public health practice, the answer to these questions means obtaining data on how healthy—or how sickly—is the population to be served. What, precisely, is meant by "levels of health"? Are we to be content with traditional, "negative" health measures in terms of morbidity and mortality? Is there a set of "markers" which can be measured in a population and which would enable one to *predict* the vulnerability of persons to specific diseases and/or to ill health? How can one get valid and reliable information on agreed-on measures of health status? Can this be done in such a way as to allow application of the results to appropriate sub-groups in the population? Can methods of assessment be developed which are economical enough to be within the resources of developing nations and which are amenable to frequent up-dating?

Among measures presently taken to help provide answers to these questions are the

following: mortality registration, morbidity reporting (both for specific diseases and on the basis of disability, absence from work, hospital admissions, etc.) case registers, physical examinations of sample populations (often using standardized and "short-cut" methods), household survey methods, "well-person" screening examinations, and special population-based diagnostic surveys. All of these methods, taken individually, could profit from methodological refinement, and there are no doubt many new innovations which could be developed. A whole area of epidemiological investigation thus falls four-square within the area of public health practice research.

2. *The development and testing of new methods and programmes of health service delivery*

Here the problems are: given a relevant medical technology, a specified set of population needs, and an adequately diagnosed social situation, what is the theoretically optimal method of providing health care services? Will this give the theoretically predicted results when put into practice?

In this, as in the case of most situations where innovations are at issue, a primary problem is to escape from the constraints of present thinking and to relate needs to the full range of potential technical solutions thereto. There is a tendency, certainly aided and abetted by vested interests and by desires for continuity and security, to stick to contemporary solutions; this is particularly true when modes of deployment and organization of personnel are at issue. A major problem for research in this area is how to assure that a full range of alternatives can be generated, evaluated, and tested.

The problem of developing new methods and programmes is at the very centre of the process of applied science. It involves the linkage of relevant basic science knowledge and technological capability with both the specific context of the problem at hand and the range of potential means of organizing requisite resources. Basic physical, chemical and biological science knowledge, engineering skill, diagnoses of environmental situation, and sound organizational inventiveness must be combined if the search for viable alternatives is to be maximized.

New methods or programmes, moreover, must be acceptable in order to be successfully implemented. However technically feasible they may be, it is an empirical question whether or not these methods will be accepted by the population which they are intended to help, or by the professional health workers and organizations which must implement them. A new procedure or programme is an innovation and, as such, is subject to a number of general principles of social change [7]; for example, they should be compatible with the cultural expectations of those whom they affect, should not involve inadvertent changes of established social relationships, and should meet a felt need for change if motivation of acceptance is to be high. Acceptance here is seen not only as a function of target populations but of implementing organizations and personnel as well and even of those groups not directly related to the programme. The ascertainment of conditions of acceptance of new programmes and the development of means to enhance acceptability are important problems for public health practice research.

Operational testing of new methods and programmes is, needless to say, important. Such pilot testing can be seen as not only an administrative technological mechanism for "debugging" it and for establishing parameters of cost and operational feasibility [8], but also as a means for developing new knowledge about the process of acceptance of innovations, about techniques of organization and evaluation, and about the omnipresent unforeseen consequences of alteration of the stage of one part of a system upon its overall equilibrium.

We know too little, for example, of the way new programmes in public health may effect the operation of the larger systems of which they are a part. For example, what may be the effects of an onchocerciasis control program upon (a) the need for and the effectiveness of basic health services in the area, (b) the economic potential and population balance of the area, (c) the natural ecology of the area? It is a rather recent understanding in public health that such questions must be considered as new programmes are contemplated [9].

3. *The analysis of determinants of utilization of health facilities*

Health service programmes, whether new or old, can be effective only if they reach the populations for whom they are intended. It is manifestly incorrect to assume (as is often done) that because there are so many service facilities per unit of population, nicely spaced throughout a population, that this group is actually able to benefit from these services. It is important to know what groups in a population are likely to utilize facilities and to what extent; it is usually true that the correlation between need for, and utilization of, services is poor. Only with such information can a rational scheme for provision of health services be developed.

We know that a number of variables determine the utilization of health facilities: objective need (in the form of disease or other conditions requiring service); perception by relevant persons that this need does exist; accessibility of service in terms of distance, cost, time, and acceptability of product; intervening opportunities or barriers to acceptance of services (e.g., competing services, negative community attitudes toward services, competing opportunities in other spheres, etc.); quality (effectiveness) of services rendered; satisfaction with services (i.e. evaluation by consumers of the *desirability* of the service from their viewpoint).

Moreover, these items vary substantially but often in obscure ways among sub-groups in a population. Much work is required to develop effective means of measuring variables of this kind and to assess how to estimate their relative importance. Even the units for measuring and classifying the bare facts of utilization are not always clearly defined [10].

Utilization data is important as an administrative tool for evaluating and controlling service activities. But it is even more important for the development of plans for improvement and extension of services. Finally, the study of determinants of utilization is likely to prove an important means for improving our overall understanding of the functions of health services and their place in the communities which they serve.

4. *The evaluation of programmes and systems*

It has been rightly observed that the activity of evaluation should be part of administrative routine in a well-run organization; such a routine process is not within the scope of research [11]. However, the process of scientific evaluation is often problematic [12], and there is need for research in the methodological area. Particularly important is the problem of developing more sensitive measures of the relative effectiveness of health services. Although excellent designs for such research have evolved in the fields of clinical testing of pharmacologic agents, clinical field trials of vaccines, etc. [13], such methods have not yet frequently been applied to the effectiveness of whole programmes. One of the difficulties, of course, is the problem of specifying precisely what should be the target outcomes against which success or failure should be measured. Another is to create a situation controlled enough that the effects of a specific programme can be separated from other, incidental effects. A

third problem concerns the difficulty and expense of maintaining follow-up over the period which may be necessary to demonstrate programme effects.

Perhaps the principal impediment to the implementation of good, scientific evaluation, however, is beyond the sphere of scientific technique. It lies in the reluctance of persons and organizations to face the criticism which may be implied in impartial scientific evaluation. Too often the attitude seems to be "if we must have evaluation, let us be sure that it can't hurt us". One of the best ways to assure this is to conduct evaluations upon the basis of slipshod methodology and imprecise measurements; their product will perforce be either ambiguous or easily refutable. In this connection, it should be stressed that when an evaluation is conducted it may lead to further research questions of great importance to public health practice. Why did a programme "succeed" or "fail"? What factors accounted for these outcomes? To ask such questions in a scientific way can often lead to the uncovering of new insights of real importance for better practice.

5. *Sociological and economic analysis of the operation of health care systems*

Aside from the questions of programme design, of studies of utilization, and evaluation, there is much to be learned about how organizations actually work and function. Every organized effort to protect and enhance the health of a community soon becomes a *social system* in its own right—and as administrators know too well, these systems must be reckoned with. They tend not to disappear when their task is completed, they develop powerful claims upon budgetary support which stem less from their accomplishment than their existence, they acquire "spheres of influence" through which they can impede or promote specific policies, and they engender very strong and significant attitudes and motivations both among their members and those brought into contact with them. To understand the constraints—and the potentialities—inherent in any organized effort to accomplish a health goal, it is necessary not only to understand its technical and physical resources and potentialities, but also its character as a social system [14]. Administrative and sociological studies of the way organizations function are of particular importance for understanding such matters as staff or employee satisfaction, productivity and turnover, problems of coordination, community attitudes toward programmes, and (above all) unanticipated consequences of policy decisions.

Health organizations typically have certain characteristics which would seem to make analyses of social process particularly important. They are comprised of a number of different types of professional workers, each of which is expected to have a marked independence of judgement and action, but whose efforts require close coordination if health goals are to be achieved. They face the difficult problem of integrating precise scientific and technological imperatives with the wishes and needs of a human population upon which they are to be carried out. Moreover, they deal with patients or publics whom they must convince voluntarily to accept their advice; unfortunately, these patients do not always share the goals of the health service for them and rarely understand fully the rationale underlying the advice they receive. Many health units deal with areas of "life and death" which evoke powerful emotional responses in the course of daily routine, and about which deepset values are held; these emotions and values may often conflict with scientific considerations. Health services often find themselves in a relatively deprived situation in terms of economic resources despite the fact that their work may be considered of highest social and moral importance. For these and many other reasons, social and psychological factors may

"warp" the functioning of health organizations away from that which might be expected in a rational model, and may do so more than in many other kinds of groups.

Economic analysis of health care systems surely needs little justification, particularly in a field where costs loom so large and needs are so great. It may perhaps be enough to mention three kinds of problem areas, all of which demand further investigation at both methodological and substantive levels. There is need for further methodological work in the field of cost accounting, particularly under the conditions of developing countries and in situations where output units are hard to define. Cost-benefit analysis, which would not only contribute to the evaluation of specific programmes and services but would also shed light on the relation of the whole health field to national economies, are both important and badly in need of improved and simplified methods. Finally, analyses of the financing of medical care and the effect of different methods of financing upon utilization of services are important, particularly in countries where direct payment by consumers for services accounts for an important proportion of health care expenditure.

Naturally, both economic and sociological studies of health organizations must somehow be integrated if a total picture of health system process is to be obtained.

6. *Analysis of problems involved in providing the requisite components of health service systems and programmes*

There is a wide range of requisites for the operation and maintenance of public health programmes. Among these are: physical facilities of many varieties (e.g. hospitals, health centres, laboratories), scientific and technical equipment (e.g. X-ray machines, microscopes, operating room equipment), consumable supplies, and—most important—trained manpower. One set of problems for research centres about the problem of logistics: what is the best means for securing the proper resources, getting them to the proper places at the proper time. A second set of problems concerns the recruitment, selection, training, and deployment of health personnel. Here, sociological problems abound. A third set of problems facing the developing countries, though not strictly of a research sort, concerns how best to build the necessary infrastructure of training institutions, industrial suppliers, and social welfare agencies necessary to support health service institutions. In this area of analysis, the relationship between the health system and other sectors of the society becomes a particularly pertinent problem.

7. *The allocation of resources for delivery of medical care services—health planning*

Given an ongoing system, data on health levels of the population, some idea of possible methods of dealing with the improvement of health, data on utilization, effectiveness, efficiency, and functioning of the system, and some knowledge of the potential availability of requisites for health services, one must ask "how best can available resources be allocated and channelled so as to gain the most effective application of medical skills to the health problems of the population at the least cost?" This is the problem of health planning. If one believes it is answerable, it is on the assumptions that the health services really do constitute an integratable system, that a rational "best way" to organize available resources can be found, and that this "best way" can be implemented through appropriate management methods. Results from the areas of public health practice research mentioned above should go far toward providing the ingredients of an "ideal" health plan. There is need, however, to do a substantial amount of work in clarifying, from the planning and resource allocation

point of view, precisely what kinds of data in these areas are necessary or desirable. Especially in developing countries, limitations on resources demand that research be, wherever possible, harnessed to action-needs. A well-defined methodology of planning could do this. Unfortunately, at the present time none exists in the health field [15].

From many points of view, then, the problem of planning can represent a synthesis of all the research areas listed above. For planning represents an analysis of the entire *system* of health care, geared to defining how it can be made to operate more effectively and efficiently. Needless to say, in a dynamic and changing world, no plan can foresee the changes which will occur even in the proximate future. Every planning enterprise therefore must involve a mechanism for constant feed-back concerning both the changing needs and constraints of population and environment. In this sense, an effective planning organization would need to guide the research effort of public health practice, orienting it to the changing intelligence and developmental needs of the health care system.

But there is a sense in which research in the health services should also stand outside the planning enterprise. Like every administrative mechanism, planning organizations become caught up in the necessities of the present—even if this is to develop a plan for the future!—and hence become unable to make the detached appraisals which are necessary for acquiring new insight. They may, in the need to develop better operational intelligence, overlook the necessity to strike out in new substantive areas and to explore theoretical issues, the resolution of which may not appear to yield immediate "pay-off" in the practical sphere. It is the contention of this paper that these needs are no less important in an area like public health practice research than in the basic sciences.

SOME CONCEPTUAL ISSUES IN INTERDISCIPLINARY RESEARCH IN PUBLIC HEALTH PRACTICE

It is at the point where the word "theoretical" is raised that one thinks about conceptual issues. Too often, so long as the problem is seen in purely "practical" or "applied" terms, the issue does not arise. Yet it is true that there is a constant interaction between theory and practice; almost every practical problem is rooted in a theoretical issue, and every theory, if well formulated, should have implications for application to the real world [16]. There are, however, some issues which appear to this author to be of fundamental importance in the field of public health practice research because their resolution tends to determine to a high degree the choice of methodology which may be employed. I shall briefly discuss six of these.

(a) *What is the nature of a system?*

There seems to be a high degree of consensus that health services constitute an analytical system. This is to say that they are composed of a set of elements which stand in relationship to one another and which together seem to affect and be affected by the environment in some coherent way. Insofar as all organizations—and groups of organizations—are systems, it follows that the field of public health practice must involve systemic elements. In the present discussion, however, I shall not attempt to show that public health services represent a system, but rather to inquire briefly concerning the implications of the concept "system" itself.

Alvin Gouldner [17], speaking of organizational analysis, has distinguished between two different representations of what a "system" is. On the one hand, systems may be conceptualized as *rational*, in that they can be described as consisting of a set of definable elements whose relationships one to another (and to the environment) can be defined precisely in a set of

functional propositions. Moreover, these elements and propositions are seen as potentially modifiable within some set of defined constraints. The ideal of a "rational system" is, obviously, a mathematical system. It is abstract, logical, and subject to rational manipulation within the bounds of the postulates and operating rules of its governing logic. By contrast, systems have also been conceived of as *natural*, in the sense that empirical observation indicates that in nature there are definable elements which are seen to be in interaction with one another and which both affect and are affected by their environment. These systems are concrete and (though one's scientific faith is that they are subject to the logic of discoverable scientific laws and propositions) they present themselves as integral wholes which are not only greater than the sum of their parts, but which function in terms of an equilibrium or homeostasis which will brook outside manipulation only to a limited degree. The biological *organism* is the classical prototype of natural systems; another application of this concept is that of the ecosystem in biological ecology. Gouldner points out that those who have analysed and described organizations have explicitly or implicitly assumed either the rational or the natural model; it would appear that this is equally true in the area of public health practice.

What are some of the implications of assuming a "rational" model of a system involving human actors? First, it may be suggested that they will be seen as operating in specific, clearly defined ways and as motivated only by the relationships defined in the system. Because of this specificity, it will likely be thought possible to develop measures of the elements of the system and their performance. In rational system models, there is a tendency, moreover, toward specifying one or a few relationships as defining sufficiently the state of each element of the system; mathematical capabilities tend to set such limits, and in the physical sciences such a modelling often is an adequate representation of reality.

When such a model is applied to a purposive organization, emphasis is naturally placed upon the relation of each element of the system to the organization's objective. Each element's activity is defined in terms of its contribution to that end. The model may approach that of a machine, in which each part, in concert with all the others, contributes its bit to the achievement of the machine's function. Analysts who apply a rational systems model to organizational behaviour usually emphasize the need for flexibility and for constantly modifying the nature of the elements of the system and their relationships in order to attain maximum effectiveness and efficiency. The rational model tends to lead to an emphasis on the possibility and importance of rational manipulation of elements of the system.

On the other hand, those who use a "natural systems" model emphasize not so much the specificity of elements of the system, and their relationships, but rather the integral and holistic nature of the system itself. They see every system as emphatically as being more than the sum of the parts. And they tend to emphasize the fact that the system is in a state of equilibrium such that any change is apt to ramify throughout the whole system. Proponents of the "natural systems" approach tend most often to see their systems as *open*, and hence not subject to a rigorous determination if its state only in terms of the measurement of its characteristics; by contrast, proponents of the "rational" systems approach may often prefer to work with closed systems. A result of these characteristics is that the proponent of a "natural systems" approach may frequently tend to stress the "unanticipated consequences" of social action [18]. He will also tend to emphasize the necessity of a thorough understanding of all the systemic relationships and their implications for a proposed change affecting the system. The proponent of the "natural systems" approach may thus tend to be conservative, seemingly content to marvel at the "wondrous

equilibrium" of nature; he may feel, however, that the proponent of the "rational systems" approach is both too simplistic in his analysis and too ready to propose organizational changes without fully appreciating their consequences [19]. In the field of organizational analysis, industrial engineers and operational researchers tend to take a "rational systems" approach, while anthropologists and sociologists frequently find the "natural system" more compatible with their style of work.

What is to be emphasized is that (at least insofar as systems involving human action are concerned) neither approach is fully valid. Both are ideal types [20], and hence are important for perceiving *tendencies* in the work of various analysts and as a means of making clear presuppositions which may only be implicit. In point of fact, the "rational systems" model is potentially capable of expressing the complexities of relationship which the "natural systems" approach is concerned to protect. The latter is, perhaps, a first approximation to the nature of a complex social or biological system. More precisely, it may represent the first stage of awareness of the fact that nature and human behaviour are neither purely atomistic nor capable of being described in purely mechanistic terms. On the other hand, the "rational systems" model, because of its apparent rigour, precision and mathematical manipulability is a very attractive way of looking at social phenomena; it bears also the prestige of classical physics. But when applied to social phenomena it often involves so many simplifying assumptions as no longer to reflect reality; the systems of classical economics and Taylor's school of scientific management are cases in point.

For the researcher in public health practice this discussion may have importance insofar as he may tend to espouse implicitly one or the other of the approaches here discussed. In point of fact, it is necessary to develop rigorous "rational models" as a part of the scientific process. It is necessary also to gather empirical data on the functioning of the health care system as it is, and with full appreciation of its complexity. The products of these researches are complementary. Both enrich one another. The road to better public health practice lies not in dogma about the nature of systems or about the methods which are corollary to these, but in an appreciation of the true usefulness of an interdisciplinary approach in the field.

(b) *Biological and social bases of public health*

It is clear that in order to improve the level of health of a human population one must have understanding of the biological mechanisms which produce disease and which control the working of the human body. It is equally clear that the practitioner of public health, as one who must motivate the actions of populations and provide services through organized groups of professional and non-professional workers, always operates within a social context. It is also clear that the incidence of disease is also affected by social differences (which go far to determine the physical environment of an individual) as well as by the natural environment and by constitutional factors. One need not argue any longer about the fact that both the biological and social contexts are basic to public health practice.

The problem, from the perspective of research, is how to develop a truly interdisciplinary approach to public health practice research. Disciplinary status considerations aside, it is not easy to coordinate the two analytic levels. On the one hand, the tendency is to regard biological (and other natural science) characteristics as the only really crucial variables, assuming that when clear scientific directives on how to control a disease are available, their implementation is purely a technical problem. On the other, it is possible to take these biomedical "directives" as given and dwell exclusively upon the social, psychological,

economic, and operational problems of carrying them out. But real life is not so easily categorizable. Changing social patterns affect the biological environment, and men's psychological reactions to social situations have physiological (and sometimes pathological) consequences. On the other hand, the social organization of health services and the expectation of publics concerning their health are, willy-nilly, shaped by both the scientific knowledge of medicine and the changing conditions of the biological environment. In a word, biological and social factors *interact* to produce given health levels and to determine the human response to illness at both individual and collective levels. A basic problem for public health practice research is how to develop a framework which will take this interaction into account. This is not to say that research undertaken from the perspective of the social science or biological science disciplines alone will not continue to contribute to public health practice research. They will, of course. It is to assert that a secure and comprehensive theory of action in public health cannot exist without this integration—and at this writing it hardly exists at more than the intuitive level. It is to, say, also that the limitations and biases which come from examining problems wholly or predominantly in terms of the one perspective or the other, should be carefully allowed for in making decisions in public health practice. It is to assert, finally, that the presence of an interdisciplinary team in public health practice cannot be an efficient means of securing a comprehensive approach to problems in the absence of a common framework of scientific discourse.

(c) *How general can the principles and findings of public health practice research be?*

Insofar as it is possible to specify, clearly, environmental variables and relationships, and insofar as "human nature" is universally the same, the scientific findings of public health practice research ought to have universal validity. But human environments differ widely and the specification of relevant environmental variables is not always feasible; often health data do not go beyond demonstrating differences in the incidence of disease, or utilization of a service, by geographic area, age group, social class or other crude category. Similarly, human social and cultural settings vary so widely that it is not possible to assume that any two social structures are really comparable or that the expectations, attitudes, or motivations of any particular category of person will be the same as those of others in his group. An antinomy thus arises between the tendency to assume that a finding made in one situation will likely be applicable elsewhere and the opposite tendency to assume that all is relative and generalization dangerous. It is perhaps true that both administrators and natural scientists tend to emphasize the universal applicability of findings, while social scientists and many clinically oriented practitioners may tend to stress the idiosyncracy of given situations.

In public health practice, a common tendency is to assume that patterns of organization and programmes which are "scientifically grounded" and have been successfully implemented in one country can, with minimal modification, be implemented in others. In practice, this has often meant that organizational patterns and programmes developed in Western countries have been prescribed for all the rest. To a substantial degree this prescription has been followed, often with useful results. In other cases, the attempt to transplant programmes and methods has either failed or produced indifferent results, and one reason frequently adduced is that cultural and environmental differences are so great that different solutions from those developed in the West are necessary. The extent to which these allegations are correct—and the development of criteria for predicting the range of environmental and

cultural variations within which given programmes can be expected to be effective—is a problem for empirical research.

Comparison in public health practice research is often extremely difficult if not unfeasible. In part, this is a function of the non-existence of specific types of data in various parts of the world. In part it is because of the differing and often unknown validity of seemingly comparable data; differences in the degree of accuracy and completeness in reporting in mortality statistics, for example [21], may be so great as to obscure apparent differences between groups. Moreover, some kinds of statistics cannot be produced in comparable form from country to country; for example, in nations in which there is a *laissez-faire* system of medicine, it is difficult, if not impossible, to develop statistics on utilization which are comparable to those coming from countries where administrative reporting on utilization of health services is regularly available. Inadequacies in data collection and lack of standardization of measures among countries, thus, is a major obstacle to assessing the comparability—and hence universality—of scientific findings. But comparison is also hindered by the lack of analysis of the range of differences between countries (or other units of comparison) which would provide a typology or scale of variation that would facilitate the interpretation of other differences. Too often today we have only crude and ill-defined sets of analytic differences, e.g. "developed" versus "developing" countries; "entrepôt", "industrial", and "service" cities, etc. These often do not give us enough information to explain differences which we measure and associate with them. What is needed, then, is both better *standardized* measures to facilitate comparisons, and theoretical clarification of the interpretive basis upon which these comparisons can be made. In the absence of this, the opposing tendencies to overgeneralize, and to limit generalizations, too severely, will likely persist unchecked.

(d) *The problem of need vs. demand*

In public health practice, a great deal of stress has been laid upon the theoretical importance of relating proposed action to the real "needs" of the population. By "need" here is meant demonstrable deficits in health which are potentially subject to amelioration. Adequate data on this matter is rarely available and usually very costly to obtain; moreover, even the best "screening" surveys involve serious difficulties in coverage and predictive validity [22]. And historical measures, such as the analysis of mortality rates can only point in a gross manner toward the needs of survivor populations.

Moreover, many people who are objectively in need of health services do not know this and, moreover, would not be motivated to do something about their needs even if perceived. This lack of motivation may be partly a function of the accessibility and cost of remedial services, but substantial non-response occurs even when barriers of cost and inaccessibility have seemingly been removed [23]. Because of this fact and because of the difficulty of getting adequate and extensive data on real needs of the population, decisions about the provision of health services are often predicated in terms of "demand". By this term is meant actual utilization of health services (or other overt behavioural evidence of a desire to receive services). Indirect "demand" measures are sometimes used also; they are likely to be even more inaccurate than direct measures of demand. Since, in many places some utilization measures are administratively rather easy to come by and since demand can be taken as a measure of what services are likely to be utilized in the future, these data when applied to projections of population change sometimes form the basis of planning decisions.

It is, however, clear that neither "need" or "demand" by itself is an adequate basis for

decision-making in the health area. Both kinds of measures are needed—the need data to relate objective problems to the technical means for their solution, and the discrepancy between need and demand values as an indicator of whether these measures are likely to be accepted without changes in their accessibility or in the attitude of the population. Put another way, measures of need refer to the biological basis of public health practice, and measures of demand to its economic and social context. The problem is to find ways of synthesizing the two kinds of measures; to argue about the primacy of use or the other is to ignore the fact that both are essential.

(e) *Is there one right way of meeting public health problems?*

Public health has traditionally been a *normative* discipline. It has a generally agreed-upon goal, and the function of the basic medical sciences has been to provide "answers" to specific problems of how to cure or prevent disease. In common with most applied fields, public health and medicine have tended to develop their generally accepted bodies of procedure, and departure from these have been taken to indicate "bad practice". The fact that public health practice very often takes place in the context of bureaucratic organizations, with their official policies and regulations that must be followed, has acted to intensify this tendency to look to the "one right way". The professionalization of the public health disciplines with the consequent collective judgements on what constitutes good practice (and condemnation of "unprofessional" approaches) has also worked in this direction. And insofar as the norms of public health practice are confirmed by objective evaluation as being effective means of attaining goals, the normative approach acts toward insuring that good performance will be achieved.

At the same time, it can tend to develop a situation where orthodoxy becomes an end in itself, where innovation can be inhibited, and where inflexibility can become the order of the day. Insofar as norms are subject to constant revision as new scientific discoveries are made or as conditions change, these dangers can be minimized. There is, however, always a gap between current situations and knowledge on the one hand and norms and regulations on the other. We are also dealing here with an attitude of mind which tends to see things in absolute "black and white" terms, and to reject alternative ways of doing things.

The normative approach should work best where situations are clear-cut, where knowledge is well developed, and where problems can be reduced to one or two relevant issues. It works less well where situations are unclear, where knowledge is only partial, and where problems are known to be complex. Many health problems have the former characteristics. For example, it is not difficult to establish whether or not a population is at risk of smallpox, we know how to prevent this disease and what the consequences of failure to do so may be, and we know that factors other than the existence of immunity to the disease probably have relatively little to do with protecting an exposed individual. The norm that populations—particularly those with any likelihood of being exposed to smallpox—should be vaccinated is clear-cut and unequivocal. The only issue here can be not what to do, but how best to accomplish this task. On the other hand, when a country—itself very diverse in character—is beset with a number of health problems requiring solution and an inadequacy of resources to meet these needs, the proper normative answer is far from clear. It is even questionable whether the answer is to go as far as one can with each of the accepted normative ways of handling the several health problems; it might be better to concentrate on one at the expense of others [24]. In this situation, what is needed is not "one right way", but a set of alternatives which can be weighed and from which a strategy can be evolved.

Accepted norms should not preclude the assessment of different alternatives. For example, the norm that only a qualified physician should diagnose and treat illness should not preclude one from examining other alternatives if there is a serious and irremediable shortage of doctors in a community. Likewise, the norms of scientific medicine ought not to preclude the examination of the function and possible utility of self-treatment, folk-medicine, and native practitioners as links in a health service where needs are beyond the scope of organized health agencies to provide necessary services. Finally, the fact that there are strong professional norms concerning division of labour and organization of services should not mean that alternative designs for action should not be considered. At the level of research, norms should not be allowed to act as blinders. And should alternatives, not in harmony with current norms and attitudes, promise more effective solutions, the question of how to win acceptance for these would be a fit subject for research in public health practice.

The foregoing discussion has had as its purpose to raise questions about basic assumptions and points of view often taken, perhaps implicitly, in public health practice. When some of these assumptions are brought to light and critically examined, it may be that some apparent obstacles to progress in the field will be clarified. It is also true that a discussion of such issues raises still more problems, the solution of which must be referred to further research.

SOME METHODS OF PUBLIC HEALTH PRACTICE RESEARCH

It is not possible for this paper either to describe or to provide an exhaustive list of useful methods for meeting the problems of public health practice research. Rather, I shall simply list some of the principal types of methods which can be used, discussing briefly their possible sphere of applicability. One conclusion from developing such a catalogue will be the necessity for seeing public health research as an *interdisciplinary problem*.

1. *Epidemiological assessment techniques*

This is not the place to discuss the methodology of epidemiology [25]. What should be pointed out is that through such techniques as case-finding, taking measures of various biological parameters in populations (e.g. measuring immunological status, etc.), and through the application of screening methods, the epidemiologist gains data on the health status of populations. Two types of methodological problems are of particular significance for our discussion. On the one hand there is the task of finding ways of determining accurate "numerators", i.e. complete and reliable enumerations of some biological characteristic in a population which is relevant for explaining the nature and distribution of disease. This, of course, involves both technical issues of how to secure valid and reliable biological measurements and field problems of assuring that complete coverage of relevant variables is attained. On the other hand, there is the problem of finding adequate "denominators"—determining the appropriate population which is "at risk" or to which the epidemiological measurement can be fittingly applied. This involves both problems of obtaining sufficient accurate demographic data on populations and the theoretical problem of identifying the groups of relevance to the problem at hand. Sampling problems are also often involved; for example, clinical data are available only for select groups of the population—patients—who represent an unknown but certainly unrepresentative segment of the population. Moreover, epidemiologically significant measurements of representative population samples are often difficult to obtain. Epidemiologists thus must fuse their technical knowledge of biological and medical

measurement techniques with a sophisticated knowledge of population characteristics, and apply these under what are often, at best, "rough and ready" field conditions.

2. *Record research*

Health service organizations are characterized by the fact that they usually maintain abundant records both concerning their own activities and performance and concerning the problems and characteristics of their clientele. These records are potentially useful for most of the problems faced by researchers in public health practice. As every researcher has found to his sorrow, however, this potential cannot always be actualized either because records are incompletely kept—or, more importantly, because they were not designed to provide the demographic or evaluative detail which would make them useful to the researcher. Too often, in every organization, record-keeping becomes ritualized and thus not very useful to the administrator, the practitioner or the researcher. An important problem is not only to standardize records so that comparability of statistics from organization to organization can be achieved, but also to plan them from the viewpoint of assuring that they will provide at least basic demographic and operational facts. And researchers in public health practice research need to arrive at a consensus concerning what these basic facts are and how they can be regularly obtained.

But in most cases, single records on individuals cannot be expected to yield all the information necessary to answer research questions. Individuals frequently must receive services from more than one organizational unit; this raises the problem of *record linkage* [26]—of how to relate information from one record to that on another. Often, moreover, information from records must be supplemented by gathering new data from the population itself. This involves the problem of identifying and locating these individuals in the community, an enterprise which is often made more difficult than necessary by inadequate record information. A whole methodology for follow-up and location of individuals, on the basis of record information for follow-up studies, has emerged; though often extensive to apply, it has made possible the accomplishment of seemingly almost impossible research tasks [27].

3. *Survey methods*

Both the social sciences and epidemiology depend for a great part of their empirical data on survey techniques—on the systematic gathering of information from a population by interview, questionnaire, or other means of contact [28]. A whole technology has evolved to support these needs, with a plenitude of specific devices and methods. Here it is only possible to mention four types of problems which must be solved in order to use survey methods effectively. First, there is the problem of selection of the population. This always involves both determining what population group is most relevant to accomplish the aims of the survey and the issues of *sampling*—how to produce (at reasonable cost) a set of persons (or events) to be studied which will be representative of the desired population. Second is the problem of instrument construction and validation. This refers to the design of means for producing an array of desired data which will throw light on the problem to be studied. What measures are available? How can they be combined efficiently into a single data-gathering instrument? Do the selected measures actually work and do they really measure what they were selected for? The problem of instrument construction and validation thus always involves both issues of effectiveness and efficiency; its importance and the necessity for adequate pretesting of instruments can of course not be stressed too much. Third, there is

the problem of *obtaining cooperation from the population to be studied.* How can necessary rapport be established? What can be done (if anything) to enhance the willingness of subjects to cooperate? How can the cost of cooperation on the part of the population studied be minimized (in terms of time, effort, exposure to stress, protection of privacy, etc.)? What follow-up measures can be instituted in order to assure that as many as possible of the desired population do cooperate? How can the necessary personal contacts which may be involved in survery work be so handled as to avoid introducing bias into the information received? Questions of this type have been particularly studied in relation to the techniques of interviewing [29], but apply to all types of survery methods. They indicate the importance of a sound mastery of human relations for successful survery research. Fourth, is the problem of designing a paradigm for analysis of data. This should, of course, be considered in advance of data-gathering insofar as possible; the problem almost always involves the application of techniques of statistical inference. The purpose of this summary listing of problems is to demonstrate the interdisciplinary nature of issues raised in survery methodology: statistical, demographic, psychological, and sociological issues are involved in the successful implementation of survey methods whatever may be the substantive area to which they are applied.

4. *Observational methods*

Often, neither clinical–biological measurements, survey methods nor record studies will suffice to answer the questions posed in public health practice research. It is often necessary to find ways of directly observing the performance of individuals or organizations; this is particularly true when the problem at hand involves a situation involving a high degree of interaction or where the relevant variables are known to be many and not easily specifiable.

At least three types of observational methods should be mentioned here. There is, first, the "anthropological" method [30], in which the investigator himself observes the persons or groups to be studied and develops a synthetic description of the reality he measures. His observation may either be done in a way that is as far as possible detached from the situation being studied or as a by-product of his participation in that situation; in either case, the question of how the fact of observation affects the situation must be raised [31]. And the problem of investigator bias is obviously particularly acute. Both the strength and weakness of anthropological observational methods lies in the fact that they depend upon the sensitivity and interpretative ability of a trained scientific observer for their results. They almost always provide configurational, holistic data as opposed to the specific, questionable results of survey methods.

Second, mention should be made of a variety of formal methods of observation which are characterized by providing criteria for an observer to systematically abstract bits of data from an observed field. Time and motion studies are an example from the domain of work study [32]: in the social science field, a number of formal methods of analysing social interaction have been developed; for example, "interaction process analysis" [33], a method for quantitatively assessing the verbal interaction taking place in a group. Such formal observational methods rely on the ability of observers to make judgements about whether what he sees fits the criteria imposed by the method; they all involve, therefore, the problem of observer reliability.

Third, mention should be made of the potential use of sound recording and photographic equipment to record situations for further analysis. These technological methods, however, while making possible accurate recordings of what happened, do not solve the problem of

analysis. They must be subjected to some kind of formal analytic procedure if they are to transcend pure description or illustration.

5. *Economic research methods*

Public health practice research must also utilize economic research methods in order to assess the feasibility and efficiency of its programmes. Development of standardized methods of cost-accounting in the field of health services represents a continuing need, and one which must involve both the skills of accountant and economist and the technical judgement of public health experts. The economic analysis of demand for services can throw an important light on problems of utilization. Economic analysis also can suggest answers to the problems of resource allocation. Cost-benefit analysis [34] in particular provides a means of weighing programme alternatives in terms of a measure of their relative *net* contribution to the society. As yet, unfortunately, cost-benefit analyses (and much of the area of cost-accounting as well) often involve a set of arbitrary assumptions and conventions which may make their conclusions controversial. But they offer hope of making decision among alternative course of action more clearly justifiable, in terms of rigorous and logical argument, than has yet been possible.

6. *Mathematical modelling and computer simulation techniques*

Ours is an age in which it is possible to gather vast quantities of data and in which the stress toward developing quantitative measures has produced more and more available "figures" for administrators and researchers to contemplate. Too often, of course, these figures are less accurate and precise than they seem. But equally important, they too often are not utilized, or are analysed at a low level of efficiency. Mathematical techniques are being developed which allow new and important insights to be developed from available quantified data. The development of multivariate analysis techniques in recent years has opened many new doors of immediate importance [35]. Mathematical analysis is increasingly able to reduce complex relationships to precise form; it can build mathematical models which can formally describe and analyse complex phenomena [36]. The public health practice researcher, like the scientist in every field, owes it to himself to learn of the potentialities of these mathematical methods for his problems. They can enhance the rigour of his thought and—thanks to the availability of the computer and its ability rapidly to manipulate logical materials—make it possible for him to investigate complex relationships with a degree of precision never before possible.

Mention should also be made of computer simulation [37] as a method of increasing potential. If a set of parameters can be defined and a set of relations stated which obtain among them, modern computers can quickly explore the finite range of situations that these parameters entail. Often the computer can, using empirical and probability methods, arrive at solutions beyond the range of more formal mathematical analysis. These potentialities make it possible to see what would happen if a whole set of events might occur. The analysis of hypothetical possibilities is thus tremendously enhanced, and it becomes possible to take a complex set of measures and study what would happen if they changed according to the relationships we suspect might hold. Computer simulation thus makes it possible to experiment with the possible implications of change (assuming some hypothesis as to its direction and effect in a set of known parameters). It makes it possible also to see if descriptions of the process of complex systems really approximate reality. Thus epidemics, metereological systems, military campaigns and complex economic phenomena are among the types of

phenomena upon which computer simulation has thrown light. It can also be used to make predictions of future states which can then be checked in reality. And many have suggested it as a valuable adjunct to experimentation. Though costly and time consuming, computer simulation certainly adds to the range of possibilities in quantitative science; as such, it certainly deserves the attention of the field of public health.

7. Operations research and systems analysis

In recent years, the mathematical methods which have been applied so successfully to the natural sciences and engineering have been focused upon the logistical and scheduling problems of complex organizations. The resultant body of interpretations, and techniques has come to be called operations research [38]. This field has demonstrated the utility of a number of analytical and control techniques (e.g. linear programming, PERT), for the solution of operational problems. More broadly, operations research attempts to apply scientific methods of analysis to the process of organized action; such an analysis always points in the direction of conceptualizing this in terms of man-machine systems. The practical results of operations research have been, in many cases, to demonstrate how to optimize results; the field has become an increasingly important contributor to the field of modern industry. Since health is among the ten major industries of industralized society, the implications for public health practice are obvious.

It is clear, however, that operations research conceived of as a formal and mechanistic kind of analysis, cannot provide all of the answers to the problems of complex organization. Although mathematics—and particularly the kinds of techniques utilized in operations research—are useful for the analysis of complex systems, they cannot of themselves deal with the complexities of human interaction or with the relevant natural science problems involved therein. Accordingly, a broader conception has developed, that of *systems analysis* [39], which is seen as an interdisciplinary approach to the analysis of complex organizations and programmes. Systems analysis is thus a synthetic discipline, devoted to the analysis and improvement of the operations of our technologically sophisticated work-organizations. This field is still in an early stage of development, but its philosophy and approach is of central importance to a field like that of public health practice. It is, perhaps, a more sophisticated and broad-gauged bearer of the traditional concerns of administrative or management "science".

All of the kinds of methods described above are either capable of generating large quantities of data or demand ready access to it. Collectively, they put extremely heavy demands upon the ability to process and analyse data. This serves to highlight the central place of modern data processing techniques in public health practice research as in all other scientific disciplines. This is not the place to stress the usefulness or the wonders of the electronic computer or of the various devices of preparing data for computer analysis. Rather, one must stress the fact that because these electronic methods are so central to modern analysis, much thought is required during the design of research concerning how to optimize their use. There is need to avoid, wherever possible, cumbersome and time-consuming hand-coding methods. Similarly, there is need to develop means of bringing the benefits of computer analysis into field situations so that investigators can gain early feedback on what they are doing—and to accomplish this at reasonable cost. So important is the use of electronic data processing and computer methods that specific plans for *how* they are to be used should be included in the planning of every research project. But at the same time, it is well to remember the implications of that "first law" of computers: "Garbage in, Garbage

out". Although they tremendously enhance the speed and accuracy and scope of research, they cannot affect the essential factors which determine its quality.

The foregoing paragraphs should make clear again the interdisciplinary nature of the skills involved in public health research. At the least, they involve theories and methods involved in the following disciplines: epidemiology, economics, geography, demography, sociology, biostatistics, applied mathematics, operations research and systems analysis, and psychology. Similarly, they depend upon the insights and practice skills of several professions: public health, medicine, nursing, administration, social work. The implication is that public health practice research must involve interdisciplinary *team* research in many cases. And interdisciplinary research involves many problems of human relations and of the necessity to harmonize recalcitrant points of view [40]. It implies, too, that persons involved in this must to an extent be generalists in their background; without some appreciation of the variety of disciplines involved, they will neither be able to apply the skills and information of their own special field intelligently nor contribute to the synthesis that public health practice as an area of knowledge implies.

CONCLUSIONS

The final paragraphs of this paper are intended to highlight a few problems inherent in the development of this field. The hope is that these problems will provide some focus for discussion concerning what is needed in the further development of public health practice research.

We may begin with the problem of developing and maintaining an authentic interdisciplinary approach. If the foregoing pages have succeeded in establishing anything, it is that such an approach is necessary. If so, then certain corollaries follow. Public health practice research groups should include specialists in several, if not all, of the disciplines mentioned above. None of these disciplines can be seen as having priority status; interdisciplinary research develops best when colleagues can work together in an atmosphere of openness and equality. But all of the disciplines must equally realize the limitations of this field and the tentativeness of their conclusions. And central to developing a good interdisciplinary approach is to select salient specific areas for investigation which are amenable to analysis from multiple points of view. Only so can the interdisciplinary approach come down from the clouds of competing frames of reference and begin to produce an organic body of knowledge. There is, therefore, need to develop a cadre of researchers from the various disciplines involved in public health and to train them in appreciation of the contribution of disciplines other than their own to this field. A set of career possibilities for such persons must also be evolved, else their availability cannot be secured. The direction of interdisciplinary research must be toward the solution of specific problems with the context of a growing, synthetic overall theory. Such needs place heavy demands upon public health leadership.

Public health practice research cannot be carried on without the active support and cooperation of a number of organizations. The problem of gaining and maintaining institutional rapport is therefore critical to its development. This means more than the willingness of organizations to cooperate in principle with studies. Often it may demand the willingness to contribute funds or other resources or to endure some inconvenience in daily routine. It means also the willingness to consider the implications of conclusions which may point to the need for revision of established policies, and in some cases to face the possibility that research findings may tarnish institutional images. All of this, of course, is

tolerable only if the results of research have some promise of improved public health practice. Yet at the same time, if this promise is to be fulfilled, the questions asked by researchers may sometimes seem irrelevant to practitioners and may carry no *immediate* practical "pay-off".

Researchers must therefore be allowed freedom, both to probe and to be, at times, seemingly irrelevant. All of this implies that if public health practice research is to develop, it must secure a considerable commitment to its causes from leaders in the field of public health. This commitment must be strong enough to allow a suspension of judgement until the chronically slow research process has had time to produce results. And it must be able to survive considerable political pressures, which will surely arise if the research is incisive and competent.

Viewed from a world perspective, one of the critical problems of public health research is to develop results which will be *generalizable* from one situation to another. The problems of developing frames of reference which allow comparability have already been mentioned. Here one need only add that, just as the needs of public health are world-wide, so also is the need for a framework of theory and methodology within which solutions to these needs, as they present themselves in specific forms, can be found. The implication is that no research in this field is good enough unless it considers the general implications for knowledge which may arise from its specific problem. Again, this points to the necessity not only for a high degree of scientific competence, but for a certain latitude for investigators to "look at the big picture" as well as at immediate needs.

But general theory and methods must also be adapted to the demands of specific problems and conditions. A major issue in public health practice research is the adaptation of methodology to field conditions, whether these be in developed or in developing countries. Several factors must be considered here. In some situations, theoretical and methodological sophistication simply outruns what can feasibly be undertaken in given research situations. For example, delicate laboratory determinations frequently cannot be made on materials taken from populations who live far from hospital facilities; and there is little use in trying to apply cost/benefit analysis in an operating system where accounting data are not available. Moreover, in many situations there are no fully trained technicians and technical assistants available, and this may necessitate the simplification of research procedures. Often, therefore, compromises and improvisations must be made by the researcher. There is a need for developing and codifying such "field adaptations". The experience of investigators into work in developing countries needs special study in this regard. On the sociological side, particularly, the problem of cultural differences is relevant in this connection. Problems of translation of instruments from one language to another are not insignificant. More important is the problem of the cross-cultural comparability of responses to the same kind of question or stimulus when the perceived meaning of this stimulus is variable. Yet despite the necessity for adaptation of methods to specific conditions, adjustments should not be allowed to prejudice either the generalizability or the reliability of research findings.

It goes almost without saying that the need and potentialities of public health research far outrun the resources which are available to it; this is the more serious because so much is still a potential, and so little has thus far been accomplished. We lack adequate theory and methods, resources are too meagre, and trained manpower is in desperately short supply. In fact it is likely that extensive training of new researchers will be necessary in most countries if research in public health practice is to make headway at all. Also, it is important to try to motivate established investigators in relevant disciplines to accept the challenges of this

field. But time is important; therefore, there is a need to develop "short-cut" mechanisms to meet the research needs of the future. Not only is it desirable that methods be simplified so that its costs may be reduced, but it is important to find more effective ways of communicating the results of research to practitioners—and to scientists involved in the development of theory. There is too much isolated effort in the field of public health practice research. Unnecessary duplication of effort thus takes place, and too many practitioners have no way of learning of the findings of others. A basic prerequisite to progress in this field, therefore, would seem to be attention to its professional and institutional infrastructure.

REFERENCES

1. Compare the discussion in the *Director-General's Report on the Medical Research Programme of WHO*, 1958–1963, quoted in GRUNDY, F., Consultation of Experts in Research in Public Health Practice. Background Paper, p. 3.
2. For a discussion of the meaning of these concepts, see BARNARD, CHESTER I., *The Functions of the Executive*. Harvard University Press, Cambridge, 1938.
3. See GRUNDY, *op. cit.*
4. LEAVELL, HUGH R. and GURNEY, CLARK E. *Preventive Medicine for the Doctor in His Community*, p. 10, 2nd edn. McGraw-Hill, New York, 1958.
5. The emphasis on application of scientific knowledge also would rule out both folk medicine and organized non-scientific activities having health goals.
6. This tendency is aggravated by the tendency in many countries to define the former as a public responsibility and the latter as within the private sphere.
7. See *Cultural Patterns and Technical Change* (edited by MEAD, MARGARET), A Manual prepared by the World Federation for Mental Health, UNESCO, New York, 1955); and *Human Problems in Technological Change: A Casebook* (edited by SPICER, EDWARD H.). Russell Sage Foundation, New York, 1952.
8. As will be noted below, much of this kind of work can be begun prior to field testing by an intelligent use of systems, analysis and planning methods or by use of simulation methods.
9. World Health Organization, *Assessment of Economic Impact of Onchocerciasis*, A Report of an Informal Meeting of WHO Advisers, Geneva, 21–23 March, 1968.
10. For example, is an "admission" or a "patient-day" statistic the most important measure of hospital utilization and what different kinds of information do they give? How should admissions, etc., be classified—in terms of first versus subsequent, paid versus charity, etc.?
11. See GRUNDY, *op. cit.*
12. SUCHMAN, EDWARD A. *Evaluative Research: Principles and Social Action Programs*. Russell Sage Foundation, New York, 1967.
13. POLLOK, T. M. *Trials of Prophylactic Agents for the Control of Communicable Diseases*, "WHO Monograph Series", No. 52; World Health Organization, Geneva, 1966; ARMITAGE, P. *Sequential Medical Trials*. Blackwell Scientific Publications, Oxford, 1960; and *Medical Surveys and Clinical Trials: Some Methods and Applications of Group Research in Medicine*, 2nd edn. (edited by WITTS, L. J.). Oxford University Press, London, 1964.
14. HENDERSON, L. J. Physician and patient as a social system, *New England. J. Med.* 212, 819–823, 1935; FOX, RENEE C. *Experiment Perilous*. The Free Press, Glencoe, Ill., 1959; STANTON, ALFRED H. and SCHWARTZ, MORRIS S. *Mental Hospital*. Basic Books, New York, 1954; PAUL, BENJAMIN D. *Health, Culture and Community*. Russell Sage Foundation, New York, 1955; PARSONS, TALCOTT, *The Social System*, pp. 428–479. The Free Press, Glencoe, Ill., 1951; and, for a particularly good non-medical study showing social system influences, see HUMANS, GEORGE C. *The Human Group*, pp. 48–80. Harcourt & Brace, New York, 1950.
15. *Medical Care in Developing Countries: A Symposium from Makerere* (edited by KING, MAURICE). Oxford University Press, Nairobi, 1966; and WHO Expert Committee, *National Health Planning in Developing Countries* "World Health Organization Technical Report Series", No. 350, Geneva, 1967.
16. MERTON, ROBERT K. *Social Theory and Social Structure* (rev. and enlarged edn.), pp. 85–117. The Free Press, Glencoe, Ill., 1957.
17. GOULDNER, ALVIN W. Organizational Analysis, *Sociology Today* (edited by MERTON, ROBERT K., BROOM, LEONARD and COTTRELL LEONARD S. JR.), pp. 400–428, Basic Books, New York, 1959.
18. MERTON, ROBERT K. The unanticipated consequences of purposive social action. *Am. Sociol. Rev.* 1, 894–904, 1936.
19. For this reason, those with a conservative ideology often find the "natural system" approach compatible with their wishes, while those who desire social change may find the "rational systems" model more attractive.

20. For a discussion of "ideal types" see GERTH, HANS H. and MILLS, WRIGHT C. (eds.). From WEBER, MAX: *Essays in Sociology*, esp. pp. 59–74. Oxford University Press, New York, 1958.
21. World Health Organization, Regional Office for Europe, *Studies on the Accuracy and Comparability of Statistics on Causes of Death. Final Report*, Copenhagen: WHO Document EURO-215.1/16, 8 February, 1966; and European Technical Conference on Mortality Statistics, *Mortality Statistics* "Report of a European Technical Conference sponsored by the Regional Office for Europe of the World Health Organization, Asnieres-sur-Oise, France, 23–28 October 1961", Regional Office for Europe, World Health Organization, Copenhagen, 1962.
22. COCHRANE, A. L. A medical scientist's view of mass screening. *Publ. Hlth* 81, 207–213, 1967.
23. DEASY, LEILA C. Socio-economic status and participation in the poliomyelitis vaccine trial. *Am. Sociol. Rev.* 21, 185–191, 1956; and FRIEDSON, ELIOT, *Patients' Views of Medical Practice*. Russell Sage Foundation, New York, 1961.
24. PIOT, MAURICE and SUNDARESAN, T. K. *A Linear Programme Decision Model for Tuberculosis Control: Progress Report on the First Test-Runs* "WHO/TB/Technical Information/67.55", WHO unpublished document, n.d.
25. TAYLOR, IAN and KNOWELDEN, JOHN, *Principles of Epidemiology* (2nd edn.). Little, Brown, Boston, 1964; and MACMAHON, BRIAN, PUGH, THOMAS F. and IPSEN, JOHANNES, *Epidemiologic Methods*. Little, Brown, Boston, 1960.
26. ACHESON, E. D. *Medical Record Linkage*. Published for the Nuffield Provincial Hospitals Trust by the Oxford University Press, London, 1967; and International Symposium on Record Linkage in Medicine, *Proceedings of the International Symposium on Record Linkage in Medicine*, 1967, E. & S. Livingstone, Edinburgh, 1968.
27. For example, see ROBINS, LEE, *Deviant Children Grow up: A Sociological and Psychiatric Study of Sociopathic Personality*. Williams & Wilkins, Baltimore, 1966.
28. KISH, LESLIE. *Survey Sampling*. John Wiley, New York, 1965; and SELLTIZ, CLAIRE, *et al.*, *Research Methods in Social Relations* (rev. one-vol. edn.). Holt, Rinehart & Winston, New York, 1964.
29. See for example, CANNELL, CHARLES F. and KAHN, ROBERT L. "The Collection of Data by Interviewing", *Research Methods in the Behavioural Sciences* (edited by FESTINGER, LEON and KATZ, DANIEL), pp. 326–380. Holt, Rinehart & Winston, New York, 1953.
30. See, for example, ADAMS, RICHARD N. and PREISS, JACK J. (eds.), *Human Organization Research: Field Relations and Techniques*, Published for the Society for Applied Anthropology by the Dorsey Press, Homewood, Ill., 1960.
31. This observation applies, of course, to all methods of studying human subjects to a greater or lesser degree.
32. NADLER, GERLAD, *Work Design*. McGraw-Hill, New York, 1960.
33. BALES, ROBERT F. *Interaction Process Analysis: A Method for the Study of Small Groups*. Addison-Wesley, Chicago, 1950.
34. KLARMAN, HERBERT E. Present status of cost-benefit analysis in the health field. *Am. J. Public Health* 57, (11), 1948–1953, 1967; and FELDSTEIN, M.S. *Measuring the Costs and Benefits of Health Services*, A Report prepared for the World Health Organization, Regional Office for Europe Symposium on the Efficiency of Medical Care, Copehnagen, 4–8 July, 1966.
35. COOLEY, WILLIAM W. and LOHNES, PAUL R. *Multivariate Procedures for the Behavioural Sciences*. John Wiley, New York, 1962; and BLALOCK, HUBERT M. JR. *Causal Inferences in Nonexperimental Research*. The University of North Carolina Press, Chapel Hill, North Carolina, 1961.
36. See BAILEY, NORMAN T. J. *The Mathematical Approach to Biology and Medicine*. Wiley, New York, 1967.
37. PIOT, M. A. *A Simulation Model of Case Finding and Treatment in Tuberculosis Control Programmes* "WHO/TB/Technical Information/67/53, "WHO unpublished document, n.d.; and BRØGGER, SVEND, *A Model of Tuberculosis and of Tuberculosis Control in a Community* "WHO/TB/Technical Information/43.65", WHO unpublished document, n.d.
38. ACKOFF, RUSSEL L. and SASIENI, MAURICE W. *Fundamentals of Operations Research*. Wiley, New York, 1968.
39. HITCH, CHARLES J. and MCKEAN, ROLAND N. *The Economics of Defense in the Nuclear Age*. Harvard University Press, Cambridge, 1960.
40. BRONFENBRENNER, URIE and DEVEREUX, EDWARD C. Interdisciplinary planning for team research on constructive community behavior. *Hum. Relations* 5 (2), 187–203, 1952; EATON, JOSEPH W. Social processes of professional teamwork. *Am. Sociol. Rev.* 16, 707–713, 1951; and LUSZKI, MARGARET B. *Interdisciplinary Team Research, Methods and Problems*. New York University Press, New York, 1958.

METHODOLOGICAL AND PRACTICAL CONSIDERATIONS IN EVALUATION RESEARCH

A paper prepared for
the Conference on Evaluation Research
in Health Care with
Special Application to the
Florida Division of Health

June 16-17, 1975

Lewis Bowman
Louis J. DiMento
John S. Prochera

Department of Political Science
University of South Florida

Tampa, Florida

I. INTRODUCTION

All public programs are under increasing burden to demonstrate their usefulness. "Recession politics" make legislators, executives, and bureaucrats very sensitive to costs and to results. Even without the current funding problems most public programs are experiencing, the demand for program evaluation would still increase. Decision-makers at every level, and in every branch of government, realize that funding alone is not the answer to public problems.

The continuing development and understanding of evaluation research can help provide answers about the efficiency, effectiveness, and practicality of many programs and projects. This will help decision-makers make rigorous judgements in developing priorities among programs and in deciding about the continuation or levels of support for individual programs and projects.

The purpose of this paper is to provide an overview of the state of the art of evaluation research. We discuss several basic methodological and practical problems in performing evaluation research.

After this introduction, the paper is divided into four sections. Section II is a review of the conceptual and methodological problems of evaluation research. Section III is a discussion of a number of the more widely used evaluation techniques. Section IV covers the problems of evaluation research from a practical standpoint. Section V is a summary.

II. METHODOLOGICAL AND CONCEPTUAL PROBLEMS

In this section we will deal with selected methodological aspects of evaluation research. First, we will define and place evaluation research within the context of its parent field, policy analysis. Next, we will examine the methodology of evaluation research, then present a typology of various types of evaluation research. Finally, we will discuss several of the concepts of evaluative research within a systems-based conceptual framework.

The Study of Public Policy

The object of study of the inter-disciplinary field of policy analysis is public policy. Policy is "a purposive course of action followed by an actor or set of actors in dealing with a problem or matter of concern . . . whereas . . . public policies are those policies developed by governmental bodies and officials."[1]

Policy analysis is the study of the formulation, implementation, and impact of public policy. Some of the basic questions which policy analysis seeks to answer are: Why and how is policy formulated? How is it put into action, and with what effect?

Evaluation Research

Evaluation research, as a subfield of policy analysis, is concerned primarily with implementation and impact. It is concerned with how policies are carried out and with the impact they have.

Purpose. Evaluation research has a twofold purpose. It is deeply involved in providing reliable and objective information about the costs and benefits of

various policies to decision-makers. This assists them in making decisions concerning the allocation of scarce resources. As such, evaluation research fulfills a pragmatic or utilitarian purpose. Secondly, evaluation research seeks to provide descriptions and explanations of social behavior which can be generalized beyond one specific case. At this second level of abstraction, evaluation research seeks to increase our knowledge concerning human behavior in social settings.[2]

Contrasted to "evaluation". Methodologically, evaluation research is not to be confused with evaluation. Suchman differentiates between evaluation, which is "a general social process of making judgements of worth regardless of the bases of such judgement," and evaluation (or evaluative) research. The latter is the "use of the scientific method for collecting data concerning the degree to which some specified activity achieves some desired effect."[3] Evaluation could include "seat of the pants" research based on anecdotal, fragmentary or biased evidence whereas evaluation research must rely on scientific standards of inference.[4] As Bernstein and Freeman correctly note:

> There is a . . . qualitative difference between individuals who accept the rules of science as a guiding philosophy in their work, and those to whom such rules are either unknown or unimportant.[5]

Evaluation research strives to be scientific. It stresses empiricism, objectivity, and rigorousness, but it requires much more than emphasis on these three points to be scientific. The basic goal of all scientific research, whether basic or applied, is theory. By theory, we mean a set of propositions which are inductively tenable and deductively valid. Stated somewhat differently this means a set of propositions which are not logically contradictory and are supported by empirical evidence. There is no real research without theory.[6]

Orientations, Levels of analysis, and emphases. Evaluation research may be either program oriented or project oriented. Wholey and his associates distinguish between a:

> . . . program . . . as a provision of [governmental] funds and administrative direction to accomplish a prescribed set of objectives . . . [and] a project [as] the implementation level of a program--the level where resources are used to produce an end that contributes to the objectives of the program.[7]

Evaluation research utilizes three levels of analysis. In program assessment the level of analysis is the entire program. Program project assessment compares a number of projects within the program. In project assessment, the level of analysis is an individual project within the program.[8]

Within these units, evaluation research can have three types of emphasis. It can perform procedural evaluation, which would seek to answer the question of whether the proper procedure was followed. Questions asked in this type of analysis would be: (1) Was the proper target population reached?, and (2) Was the action taken as specified?

The emphasis may also be on substance. This is often called impact analysis. Evaluation research in this area of emphasis would be on whether there was a desirable change in the target population which can be attributed to the action.

The third emphasis can be both procedural and substantive. In this emphasis it would examine whether the action was carried out as originally planned and whether there was a favorable impact in a specified population.[9]

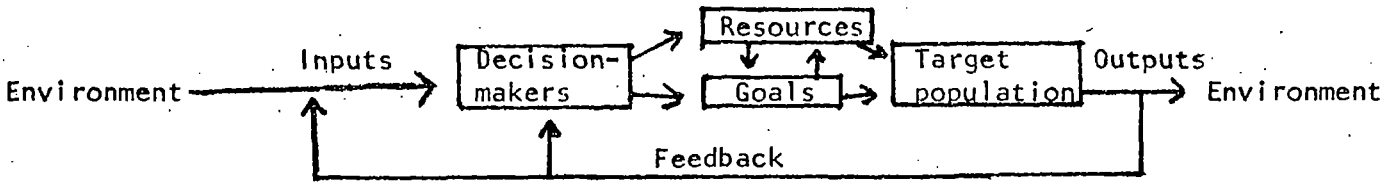
The combination of these two concepts yields the typology presented in Table 1 which can be utilized to classify the types of evaluation research, both in terms of emphasis and level of analysis. Type III evaluation research would thus be an examination of whether a project followed the procedural guidelines of the project.

Table 1. Types of Evaluation Research Organized by Level of Analysis and Type of Emphasis

Type of emphasis	Levels of analysis: Program assessment	Program project assessment	Project assessment
Procedure	I	II	III
Substance	IV	V	VI
Procedure and Substance	VII	VIII	IX

"System" concepts. Regardless of the level of analysis or the emphasis all types of evaluation research are unified in their use of certain key concepts. These are environment, inputs, decision-makers, resources, goals, target population, outputs, and feedback. Taken together, these concepts constitute an impact delivery system. The purpose of this system is to create favorable changes in a specified population. Whether favorable or unfavorable, however, the changes are the impact.

Figure 1. A Conceptual Framework for the Analysis of an Impact Delivery System



The impact delivery system is constructed as an analytic system instead of a natural system. Although an impact delivery system may exist as a natural

system, a set of variables which may naturally cohere, it is generally believed to be more fruitful to use the notion of an analytic system. By this we mean a system is nothing more than a symbolic representation of whatever variables in the real world the analyst selects as theoretically fruitful for explanation and prediction.[10]

The environment is conceptualized as the social system or society within which the impact delivery system operates. Because the impact delivery system is conceptualized as an open system, one which interacts with its environment, the framework includes inputs and outputs. Inputs are the influences the environment has on the impact delivery system through supports and demands. Outputs are the influences the system has on its environment.[11]

The decision-makers are those people who are in a position to make authoritative decisions concerning the setting of goals, the distribution of resources, and the selection of the target population. They are also the component of the impact delivery system most susceptible to the supports and demands which originate in the environment. The framework recognizes that the need of a target population is only one of many factors which the decision-makers must face. They also consider whether the general social environment generates a demand for this project and whether they have the support of the environment also.

Goals are set by the decision-makers in response to both the inputs from the societal system and the need of the target population. Unless there is a clear specification of goals evaluation research is very difficult. Suchman states that programs or projects which specify objectives or goals in the following way would be most amenable to evaluation research:

1. The content of the objectives
2. The target of the program or project

3. The time within which the changes are to take place

4. The number of objectives

5. The extent of the expected effects.[12]

For a large number of reasons this is seldom the case when goals are set for a project.

The resources which are allocated to any project are determined in part by the environment. These resources are scarce, both in terms of personnel and material. The resources allocated to one project are resources denied to another project. The allocation of resources is often less responsive to the need of the target population than to demands of the rest of society.

Resource utilization can have positive or negative consequences. The consequences are positive when they assist in the achievement of stated goals; they are negative when they hinder the achievement of stated goals.

Resources include personnel and material. Personnel resources are the people involved in the project. They are evaluated in terms of quantity, quality, and morale. If they are trained for the job and have high morale, they are considered a resource having positive consequences. If poorly trained and of low morale, they are a resource having negative consequences.

Material resources are those goods which the personnel need for the project or program. They are measured in terms of quantity and quality. Quantity refers to the question of whether there are sufficient resources to achieve the stated goals while quality refers to the appropriateness and the grade of the allocated material.

These simple questions of quantity and quality of resources do not totally explain their consequences. The organization for utilization of the resources and the geographical distribution of the resources also may determine whether resources will have positive or negative consequences.

The target population is the potential clientele of the project or program. The target population may be considered in terms of four criteria: (1) need; (2) receptiveness; (3) utilization; and (4) impact.

The measurement of need is of crucial importance to all aspects of evaluation research. The most rigorous evaluation of the impact of a project can take place only when need has been measured both reliably and validly. It is not an understatement to say that without good measures of need there is no evaluation research, only evaluation.[13]

Need may be conceptualized in terms of objective and subjective need. This objective-subjective distinction is a dimension of which decision-makers are becoming increasingly aware. In the past decision-makers have been guided solely by the concept of objective need: Do objective measures indicate the existence of need in a population? Subjective need takes into consideration whether the target population perceives itself to be in need: Do members of the target population know that their conditions can be improved?

Receptiveness refers to whether the target population will accept the project or program. Receptiveness is related not only to the design of the project or program but also to those specific cultural norms which will influence receptiveness.[14]

Utilization refers to the extent the target population uses the project or program. Quite clearly utilization is closely related not only to need, but also to the receptiveness of the target population. It should be noted, however, that the target population may be receptive to the project without utilizing it or vice-versa. Utilization also is related to the proximity of the project to the target population as well as to the way in which it is administered.

Finally, there is feedback which refers to the information the decision-makers and others receive concerning the impact. Evaluation research can be fitted within this model as an attempt to increase the feedback or information concerning social programs to the salient decision-makers and to society generally.

This conceptual framework and the definitions of its concepts provide an overview of the methodological problems in evaluation research. Next, we will consider selected techniques of evaluation research.

III. EVALUATION TECHNIQUES

In this section we will consider techniques concerning two general objectives of evaluation research. The first objective is to determine the existence and extent of impact that a program has on the target population. The second objective of evaluation research involves the determination of whether a program which has a demonstrated impact is worth the cost that will be incurred in implementing it.

Experimental and quasi-experimental research designs are useful in carrying out one objective of evaluation research--that of assessing program impact. Cost-benefit analysis and cost-effectiveness analysis are two methodologies which deal with another objective of evaluation research--the determination of whether the program is worth the costs. Assessment of program impact should precede the latter objective because cost-benefit and cost-effectiveness analyses require knowledge about the impact of the program.

Determining the Extent of Impact

As discussed above, one purpose of evaluation research is to determine the extent of impact that a program has on the target population. It is at this juncture that statistical analysis becomes essential. A great many statistical techniques are available to the researcher, and the utility of any specific technique depends upon the situation. However, one of the first activities which the analyst must perform is to design his experiment with the goal of obtaining as much information as possible at as low a cost as possible.

Designing experiments. The design of experiments is a very broad subject concerned with methods of sampling to reduce the variation in an experiment

and to maximize the amount of information available from that experiment. In spite of the simplicity of its objectives, it is a very complex subject. We will provide an overview of the subject, examine some of the relevant considerations, and provide an example of application.

Several steps are normally employed in designing an experiment.[15] The steps include:

1. Select the factors to be included in the experiment and specify the population parameters of interest.
   A factor is nothing more than an independent variable. The population parameters are the dependent variables that are being studied. The program being studied and its goals will usually predetermine the dependent and independent variables although there often is a certain amount of ambiguity in the goals which may make specification of the dependent variables rather difficult.

2. Decide how much information is desired concerning the parameters of interest.
   If the experiment involves estimation, how accurate must our estimates be? If it involves hypothesis testing, what probabilities of type I and type II errors do we need? A type I error means that we reject our hypothesis when it is true. A type II error means that we accept our hypothesis when it is false. The probabilities associated with these errors should depend upon the seriousness of their consequences.

3. Decide how to combine the variables and determine how large a sample should be taken for each combination.
   The total sample size as well as the number of replications (cases) per variable combination will effect the variability of the estimates and, therefore, the accuracy of the results. Replication also is necessary if the experimenter desires to determine whether there is any interaction among the factors. Interaction between two independent variables, A and B, occurs when the change in response of the dependent variable as a function of A is different for different levels (values) of B.

These steps can be clarified with the aid of an example. An experimenter wishes to determine the effect of four drug treatments on blood pressure of cats. How should he design his experiment and what are the factors that influence this design?

The experimenter may construct a completely randomized design. In this case the drugs are considered the only factor and this factor has four levels--i.e., the four drug treatments. He then selects four random samples of cats and applies one drug to each of the four samples. This design has the advantage that it is simple but has the disadvantage that the cats may differ in their response to any drug. To get an accurate sample, therefore, the four samples may have to be large and therefore the experiment can be costly.[16]

After the experimental design has been determined the researcher must proceed with his analysis. The type of analysis which is appropriate depends upon the experimental design which, in turn, depends upon these factors: (1) the nature of the program under study; (2) the variation among subjects; and, (3) the amount and accuracy of information required.

Quasi-experimental design. Unfortunately, the nature of the population or program under study often makes it impossible for the social scientist to utilize an experimental design. In this case one must engage in what is called "quasi-experimental" design.

Experimental design requires that samples be selected on a random basis. This simply means that once the population to be studied has been determined, each element in the population has an equal likelihood of being chosen in the sample. In practice, the researcher is often faced with situations in which this is impossible, and the results of the experiment may therefore be invalid unless special steps are taken. It is the purpose of quasi-experimental design to introduce techniques into the experiment which increase its validity despite the nonrandomness of the sampling procedure.

The validity of an experiment is threatened by the fact that the results of the experiment may not be attributable to the independent variables being

studied. Other variables which may bias the results of an experiment arise from two general sources: (1) the sample selected is not random so that the variables influencing the results are characteristics of the sample itself; (2) even if the sample is random, there may be other extraneous variables influencing the outcome which the experimenter has not accounted for. These variables may be completely exogenous to the experiment or they may be artifacts of the experiment itself such as the manner in which a treatment is applied.

By definition an experimental design precludes the first type of bias. That is, an experimental design requires that the samples be selected randomly. The second type of bias--the influence of extraneous variables--can effect both experimental and quasi-experimental designs.[17]

The most common means of dealing with both of the biases (nonrandomness and extraneous variables) is to include a control group in the study. This is a group to which the treatment under study is not applied. If the control group and the experimental groups are selected by the same process, the presence of a control group can increase the validity of an experiment, even if the samples are not random. If the final results for the control group differ from the results for the experimental groups, one can safely conclude that the results are not due to the selection process.

The presence of a control group also can increase the validity of an experiment which may be influenced by extraneous variables. If the control and experimental groups differ in their response to the experiment, one has reason to conclude that this disparity is due to the independent variables under study rather than any extraneous variables because the latter presumably have the same impact on the control and experimental groups. Since extraneous

variables can affect true experimental designs, a control group can often be fruitfully employed in such a design. For example, a placebo may be administered as a control drug in an experiment to test the effects of certain drugs on a population. A control group also is often utilized in a true experimental design in order to measure the dependent variable in the absence of any treatment--that is, for comparative purposes.

The very nature of the programs and activities which the social sciences study often make it impossible to take a random sample. Social scientists are often precluded from taking a random sample for political reasons or for practical reasons. The political problem may be demonstrated by this example. Knowledge of the effect which a new drug has on the incidence of heart disease is desired, and it is known that the drug has no harmful side effects and that it may be very beneficial, but the drug is available for the present only in limited quantities. The most obvious approach to this problem would be to construct a completely randomized design. A sample may be selected, the drug administered and the incidence of heart disease measured. A control group to whom a placebo is administered is necessary for comparative purposes and to account for any effects of extraneous variables.[18]

In this case, however, political problems stand in the way of this true experimental design. The law requires that the drug be administered on a first come-first serve basis. This selection process makes evaluation of the drug very difficult, and it certainly decreases the validity of the experiment. The obvious reason for this is that persons who take advantage of this drug are likely to participate in other activities which may have a substantial impact on their health.

As mentioned above, a control group is often utilized in quasi-experimental

design to counteract the effects of a nonrandom sample. However, in this case, the application of a placebo to some of the individuals who participate in the program is also legally and ethically unacceptable. Thus, political constraints, the very factor which necessitated a control group, also has made utilization of such a group impossible.

A practical problem occurs when the activity under study itself predicates the manner in which individuals are selected to participate in that activity. For example, if a researcher is studying the effect of taking Latin on a person's vocabulary he could not take a random sample of students and force them to take the course. Therefore, he is faced with the problem that students who elect to take Latin may vary appreciably in certain relevant variables from those who do not.

Measuring the dependent variable before administering the program is another common means of counteracting the two threats to the validity of a quasi-experimental design. This procedure is referred to as a pretest. Because the subjects in a quasi-experimental design are selected nonrandomly, it is possible that they may differ in their response before the program is administered. For example, students who plan to take a course in Latin may have a better vocabulary before they take the course because of their general motivation. In general, a pretest can be administered to the groups being studied and if there is no difference among these groups, the experimenter can more safely conclude that the post-test differences are due to the effect of the program under consideration.

The dependent variable also may be measured before the test is administered in a true experimental design, but such a measurement is made for reasons other counteracting the problem of nonrandom selection. For example, in an experiment

to measure the effect of a program on reading speed, the experimenter must have a measure of the reading speed in the absence of the program, that is, one must have a control group with which to compare his experimental group. In this case, the experimenter may simply measure the dependent variable for the experimental group before the test is administered. The measurement is made simply for comparison purposes and therefore may be termed a "pretest control". This comparison could instead be made on another random group to which the program is not administered.

In some situations the pretest control is inappropriate and a second group is essential. For example, the heart disease study discussed earlier probably could not have utilized a pretest control because of the extended period over which the dependent variable is measured. The experimental group would necessarily be much younger and therefore less prone to heart disease during the period over which the pretest measurements would have to be made.

A pretest, therefore, is utilized in a quasi-experimental design to counteract the effects of a nonrandom sample while it may be utilized in an experimental design as a control group to compare with the experimental group. The pretest does not alleviate the problem of extraneous variables that may affect the population over the period of the study.

Examples of quasi-experimental and experimental designs. A large number of different designs can be constructed depending upon the combination of control groups and pretests which are utilized. Some of the common combinations are listed below. In each case, if the sample is chosen randomly the design is experimental, if the sample is not chosen randomly, the design is quasi-experimental. The conventional notation which is utilized in this listing is as follows: (1) "X" refers to the experimental treatment (independent

variable; (2) "O" refers to the observation on the dependent variable; and (3) temporal order is from left to right.

I. One Shot Case Study (no control, no pretest)

X O

A. Random sample:

1. Advantages:
   a. The design is relatively inexpensive

2. Disadvantages:
   a. Baseline data on the population under study is not available; therefore, there is no way of knowing the value of the dependent variable in the absence of the treatment and comparisons cannot be made
   b. Does not observe the possible effects of extraneous variables

B. Nonrandom sample:

1. Advantages:
   a. Same as random sample

2. Disadvantages:
   a. Same as random sample
   b. Provides no way of ascertaining whether the results are due to the selection process instead of the independent variables

II. One Group, Pretest/Post-test Design (pretest but no control)

O_1 X O_2

A. Random sample:

1. Advantages:
   a. Allows the experimenter to compare the dependent variable after treatment with the dependent variable before treatment

2. Disadvantages:
   a. Does not counteract the possible effects of extraneous variables

B. Nonrandom sample:

1. Advantages:
   a. Same as random sample

2. Disadvantages:
   a. Same as random sample
   b. Provides no way of ascertaining whether the results are due to the selection process

III. Post-test Only, One Control Group Design (control, but no pretest)

X O_1

O_1'

A. Random sample: (Note this is an example of a completely randomized design which was discussed earlier.)

1. Advantages:
   a. Allows the experimenter to compare the dependent variable with and without treatment
   b. Counteracts the effects of extraneous variables since experimental and control groups are presumably both subject to these variables

2. Disadvantages:
   a. None

B. Nonrandom sample: (This is called a Static Group Comparison)

1. Advantages:
   a. Same as random sample

2. Disadvantages:
   a. Provides no way of ascertaining whether the results are due to the selection process (the control and experimental groups may have exhibited the same differences before and after the experiments)

IV. Pretest/Post-test, Control Group Design (control and pretest)

O_1 X O_2

O_1' O_2'

A. Random sample:

1. Advantages:
   a. Allows the experimenter to compare the dependent variable with and without treatment both with the pretest and control

2. Disadvantages:
   a. Since the sample is chosen randomly O_1 and O_1' are unnecessary, therefore the design is impractical
   b. The pretest itself might affect the results

B. Nonrandom sample:

1. Advantages:
   a. If there is no difference between O_1 and O_1' we can be more confident that the results are not due to the selection process. If there is no difference between O_1' and O_2' we can be more confident that results are not due to extraneous variables

2. Disadvantages:
   a. The pretest itself might affect the results

V. Four Group Design

O_1 X O_2

O_1' O_2'

X O_2''

O_2'''

A. Random sample: (This is referred to as the Solomon Four Group Design)

1. Advantages:
   a. Represents an advantage over IV, A, if it is feared that the pretest measurement may by itself affect the results

2. Disadvantages:
   a. Is unnecessarily costly since III, A accomplishes the same ends

B. Nonrandom sample:

1. Advantages:
   a. Same as random sample but in this case the advantage may be crucial in a situation where the pretest measurement may affect the results since method III, B may not provide valid results when the sample is nonrandom

2. Disadvantages:
   a. Design is costly

VI. Time Series Design (no control--a number of pretests and post-tests applied over an extended time interval)

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

A. Random sample:

1. Advantages:
   a. Same as II, A above
   b. The several pretests and post-tests over an extended period of time may assist the experimenter in identifying extraneous variables that may affect his results

2. Disadvantages:
   a. Same as II, A above although in this case the disadvantage may not be as serious since the experimenter has attempted to identify some of the extraneous variables
   b. The design is costly since III, A above counteracts the effects of extraneous variables

B. Nonrandom sample:

1. Advantages:
   a. Same as random sample

2. Disadvantages:
   a. Same as II, A above

VII. Multiple Time-Series Design (control and a number of pretests and post-tests applied over an extended time interval)

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

O_1' O_2' O_3' O_4' O_5' O_6' O_7' O_8'

A. Random sample:

1. Advantages:
   a. Same as IV, A above
   b. Also, the several pre- and post-tests over an extended period of time may assist the experimenter in identifying extraneous variables that may effect his results

2. Disadvantages:
   a. Same as VI, A above

B. Nonrandom sample:

1. Advantages:
   a. Same as IV, A above with the addition that the experimenter is more capable of identifying the presence of extraneous variables

2. Disadvantages:
   a. Costly

VIII. Multiple Program or Project Comparison Design

O_1 X_1 O_2

O_1' X_2 O_2'

O_1'' X_3 O_2''

O_1''' O_2'''

This is a means of comparing the effects of a number of programs rather than determining whether a single program has an effect. If the samples are chosen at random, O_1, O_1', O_1'', and O_1''' would not be necessary and the design then would be a completely randomized design which was discussed in III, A and earlier in the paper. If the samples are not chosen at random this design would have the same advantages as IV, B above.

As the above listing indicates, the variations in the application of pretests, post-tests and the utilization of control groups can lead to a large number of experimental designs. However, many of these variations are obviated when the samples are randomly selected, so that design III (design VIII if a number of programs are being compared) is usually the most efficient of the ones discussed above if a true experimental design can be constructed. Designs III and VIII (without pretests) are completely randomized designs.

If the samples cannot be randomly selected--that is, if the experimenter must resort to a quasi-experimental design--the choice of the above designs depends upon the circumstances. Design IV, the pretest/ post-test, control groups design or design VII, the multiple time-series design, are generally the most reliable and when feasible should be considered in any evaluation effort.

Comparing Costs and Worth

After determining the program or project's impact, the evaluators then must determine whether the worth of the program justifies its costs. One way of making this determination is cost-effectiveness analysis which specifies criteria for determining the effectiveness of a program and attempts to measure all the costs that the program incurs. Cost-benefit analysis can be considered a subset of cost-effectiveness analysis in that the effectiveness of the program is measured in monetary terms, and thus denoted as benefits.

Cost-effectiveness analysis. Whether effectiveness is measured in monetary or nonmonetary terms, certain arbitrary decisions must be made in order to define it. If a nonmonetary measure of effectiveness of chosen the researcher is faced with the problem of weighing the various impacts of a program in order to derive an overall measure of effectiveness. Similarly, in order to compare one program with another he must make judgements concerning the disparate effects

of the programs. For example, how does one compare the benefits of diabetes treatment with those of a cancer operation?

One approach is to empirically determine the utility that a sample of individuals feel they derive from various health states.[19] The average utility of the sample is then used to construct a health index. The index is constructed by asking the respondents to choose among various alternatives that may result from the program and utilizing certain mathematical techniques to rate the health states. This health index, then, is a measure of the utility society derives from the various health states that a specific program may attain. The effectiveness of a program in any year is the product of the increase in utility associated with the change in health state which it causes multiplied by the length of time the program is effective. The overall effectiveness of the program is the present discounted value of the program over the period of its operation.

This model requires that utilities be measured each time the concept is applied to a specific program because an infinite number of health states would generally have to be considered. It also requires that the changes produced by the programs under consideration in the health state of a population be determinable. Furthermore, it may be difficult for respondents to determine their preferences and they may simply resort to guessing. As a result of these and other difficulties some authors have rejected such an approach until these difficulties can be resolved by further research.[20]

Determination of the costs of a program also can be a complex process. Some factors to be considered when measuring costs will be discussed below when cost-benefit analysis is described.

In spite of these problems it is useful to discuss briefly how cost-

effectiveness analysis can be applied if (or when) a suitable measure of effectiveness can be constructed. The first consideration is that the inputs (costs) and outputs (effectiveness) are measured in different units. Therefore, unless the costs of a program exceed budgetary constraints or the effectiveness is negative, no decision concerning one program can be made. This methodology can only be used to compare a number of programs.

Two general methods of choosing among programs are: (1) integer programming (a type of linear programming) which is especially useful when there are constraints in addition to cost constraints; and, (2) a cost-effectiveness ranking algorithm .[21] For programs which are not mutually exclusive this algorithm simply states that programs with negative costs should be chosen first and top priority should be given to programs with the most negative costs. According to Torrance,[22] a program will have negative costs if the health care costs and earnings lost because of the disease in question exceed the application costs and the earnings lost owing to program participation. Positive cost programs are ranked in descending order of their effectiveness-cost ratio. Priorities among programs then are assigned on the basis of this ranking.

Cost-benefit analysis. As mentioned above, cost-benefit analysis can be considered as a subset of cost-effectiveness analysis in which the effectiveness as well as the costs of the program are measured in monetary units. Some authors reject cost-benefit analysis in evaluating health programs because the technique measures the benefits of a program in terms of the increased earnings that will accrue to those who participate. They believe this approach "fails to define many types of benefits such as the dollar value of improved health for housewives, retired people, and others with no earnings,"[23] as well as the dollar values of increased patient comfort and reduction in mental anguish.

However, the algorithm described above creates a similar problem for these authors' cost-effectiveness model. This is because the negative costs that result from a reduction in lost earnings are lower (more negative) for a thirty-five year old laborer than for an elderly widow. Therefore, if program A and program B are equally effective while the former assists the laborer and the latter assists the widow, program A would have higher priority. The authors' technique, therefore, seems to be faced with the major problem confronting cost-benefit analysis, the problem resulting from differential productivity. The different treatment of foregone earnings in their model seems to be merely semantic. Torrance and his associates treat the reduction in foregone earnings as a negative cost while traditional cost-benefit analysis treats it as a positive benefit.

Cost benefit's measurement of benefits in terms of increased earning power is probably its major drawback in the health field since many researchers and administrators would not accept this measure of effectiveness. However, in other areas, notably water-supply projects and other public works, the methodology has been used extensively. In certain areas of health care it may be an effective tool.

What does cost-benefit analysis involve? In brief, the technique requires assessing all the relevant costs and benefits of a program in monetary terms. Assuming that the costs do not exceed budgetary constraints, the program should not be implemented if the costs exceed the benefits, that is, if the benefit-cost ratio is less than one. The fact that costs and benefits are measured in the same units allows cost-benefit analysis to judge the value of a specific program in contrast to cost-effectiveness analysis which can only choose among a number of programs.

The description of cost-benefit analysis makes it seem very straightforward, but a myriad of difficulties confront the technique in practice. The major problem that confronts its application to health care evaluation has already been mentioned--the measurement of benefits in terms of the reduction in foregone earnings.

In addition to this problem, economists have spent a great deal of time grappling with other difficulties that arise in the application of cost-benefit analysis. Prest and Turvey[24] have described these difficulties thoroughly, including these considerations: (1) the enumeration of costs and benefits; (2) the valuation of costs and benefits; and (3) the choice of interest rate. In brief, these mean:

1. the enumeration of costs and benefits:
   The conceptual problem confronting the measurement of benefits in the health field has already been discussed. Another factor is the existence of externalities--that is, of costs and benefits which may accrue to groups other than those participating in a program. The major question here is deciding when these externalities should be taken into account.

2. valuation of costs and benefits:
   A number of questions are included under this topic. For example, the existence of monopoly conditions distorts the prices of factors so that investment decisions based on valuations of costs and benefits at market prices may not be appropriate. The importance of this factor in the health field, which is anything but competitive, cannot be underestimated.

   Another important consideration that falls under this heading is the level of unemployment. Because excess supply of an input results in market prices which overrate the social cost of using that input and because expenditure on a project will decrease unemployment, "The use of market values to ascertain direct costs and benefits of a project overrates it social costs and underestimates its total benefits."[25]

   If the program under consideration provides a public good, the market mechanism for determining the costs and benefits of a program breaks down.[26] This presents the analyst with another problem in his attempt to measure costs and benefits.

3. choice of interest rate:
   Because the benefits and costs of most programs accrue in the future, the present value of these future benefits and costs must be determined. The question here is what interest rate should be used in determining the present value. Some authors, for example, believe that governmental programs should use lower interest rates than those existing in the private sector because public programs must be concerned with the future.

This discussion of some of the difficulties that arise in the application of cost-benefit analysis is not meant to be exhaustive but merely to apprise the researcher with some of the practical problems that arise. It is obvious from the nature of these difficulties that they also confront the researcher if he utilizes cost-effectiveness analysis.

IV. PRACTICAL CONSIDERATIONS

Many of the practical problems facing evaluation researchers are as difficult to solve as the conceptual and methodological problems. These practical considerations include political, ethical, and organizational problems.

The political and ethical problems raise questions relating to problems inherent in any endeavor which utilizes, or appears to utilize, strategies of social experimentation. Bureaucratic actions and reactions toward evaluation research, and the problems posed when evaluation research is used, or perceived, as a "political tool" also raise political and ethical questions.

Organizational problems raise questions about the advisability of utilizing internal or external evaluations, about the antipathy of evaluators and administrators, and about the utilization of social scientists as part of the evaluation research team. Feedback structure for utilizing the findings, plus the funding levels which extensive evaluation research requires, raise additional organizational problems.

Questions of Politics and Ethics

The difficulties which ethical and political considerations pose for evaluation research have received wide discussion but little resolution.[27] The more common of these difficulties include the problems involved in social experimentation, the differing perceptions of need for evaluations and utilization of evaluations, and the fact that increasingly evaluation research has become a "political tool."

Problems in social experimentation. Evaluation research in social and

economic public policy areas involves considering policy outcomes in terms of their impact on human beings. These outcomes, and expectations surrounding them, are inherently controversial.[28]

Researchers in the field are only too aware of the trouble that allegations of "exploiting clients" can pose for personnel seeking to randomize services and related phenomena. Though varying widely in evaluation efforts, the problems here are the difficulties generated in studies utilizing human subjects in any research design.

Some researchers are so concerned about this problem they refuse to consider fully utilizing correct research design and methods even when they are available. For example, in their description of the evaluation of a community mental health program, the researchers explicitly cited the "exploitative and unethical"[29] nature of the experimental approach as a reason for opting for less rigorous evaluation design.

Researchers voicing these views do not seem to take into account the possibility that, ultimately, the most serious exploitation is the deliberate decision not to design the evaluation sufficiently rigorously to provide definitive or generalizable answers. A more rigorous research strategy would ensure more effective treatment for all clients over a period of time. It also would provide justification for a public policy ensuring adequate funding levels to meet the needs of many more clients. This argument notwithstanding, the controversies about the question of social engineering as one aspect of adequate evaluation research produce continuing ethical and political questions.

Bureaucratic actions and reactions. Survivors in private or public bureaucracies learn to "cover their trails well" early in career. The higher the level of appointment, the greater the necessity to pay attention to this rule. Hence,

the initial inclination of many bureaucrats is to resist evaluation generally and to resist rigorous, systematic evaluation with a vengeance. Wessen has noted:

> . . . the reluctance of persons and organizations to face the criticism which may be implied in impartial scientific evaluation. Too often the attitude seems to be 'if we must have evaluation, let us be sure that it can't hurt us.' One of the best ways to assure this is to conduct evaluations upon the basis of slipshod methodology and imprecise measurements; their product will perforce be either ambiguous or easily refutable.[30]

Administrators have need to believe their programs are successful and may actually seek ambiguity of research design to ensure ambiguity of evaluation results.[31] Often, if results are unfavorable to on-going programs or plans, the first line of attack will be on the methodology utilized in evaluation.[32]

Administrators often are reluctant to accept the results of the evaluation. Many evaluations of social action programs have shown that many programs have little immediate impact on the behavior of groups or individuals.[33] Administrators of programs naturally view such findings with alarm fearing the findings will be utilized to argue for termination of the programs. Of course, if properly understood, the problem of utilization of negative evaluation findings is much more complex and often does not necessarily lead to recommendations for program termination. For example, redefinition of program goals to make them more realistic, restricting delivery systems to make them more effective in producing the projected programmatic goals, and many other strategies usually are the proper recommendations resulting from negative evaluation findings before considering program termination.

Evaluation as a "political tool". Policy makers at every level of government are increasingly demanding "accountability" in personnel, services, and programs generally. These demands often are put in the harsh terms of immediate accountability, and, as such, are often "political" demands rather than

efforts to ensure evaluation in a strict sense.[34] In fact, this tactic is an effective scapegoat technique in periods of economic recession or depression, and as such, one turned to regularly currently. However, legislation which deliberately mandates evaluation research is not necessarily politically motivated and often serves a useful purpose. Since the middle 1960's, Federal legislation or administrative regulations have mandated evaluation research in a number of programs.[35] The same phenomenon is occuring at state levels, the action of the Florida legislature in the Health and Rehabilitative Services Reorganization Act being a recent example.

Organizational Problems

Several practical considerations are involved in deciding about the organization of evaluation research. Should the evaluation be conducted by persons internal or external to the program or project? How are the "feedback" problems to be solved to ensure the utilization of the evaluation? Is it possible to alleviate the points of contention which often arise between administrators of programs or projects and evaluators of the programs or projects?

Internal or external evaluators? Although the discussion is often in the framework of this simple dichotomy, the possibilities for variation on the theme are relatively wide. For example, the evaluation staff may be composed of persons: (1) solely in the employ of the program or project being evaluated; (2) in the employ of the program or project being evaluated and the parent project or program; (3) partially in the employ of the program or project being evaluated and an outside evaluating agency; or (4) in the sole employment of an outside evaluating agency. These variations have different mixes of the internal-external evaluator problems, however, we shall discuss only the problems and advantages of the first and the fourth.[36]

Caro has listed several arguments in favor of employing outsiders as evaluators, writing that:

> . . . (1) they tend to be better able to maintain their objectivity; (2) they are more likely to be able to include evaluative criteria that question basic organizational premises; (3) they may be able to mediate more effectively where there is extensive internal conflict; (4) they usually are better protected from problems of marginality and status incongruity; and (5) they are better able to avoid unwelcome nonresearch tasks.[37]

However, outside evaluators often undermine these advantages by their tendency:

> . . . to design studies in accord with their own interests but which may be far afield from the main policy-related questions that need to be addressed. Also . . . [they] have a tendency to overstate in their initial proposals what their studies realistically can accomplish.[38]

Internal evaluators offer the advantages of having more detailed knowledge of the agency, its programs or projects, and its objectives. Plus, if adequately funded and staffed, internal evaluators are in a better position to provide continuous evaluation research.[39] This latter advantage is tenuous, however, because evaluation funding often is the easiest mark for program administrators when they have to make choices about internal funding priorities.

External evaluators must guard against a temptation to lose their objectivity rather than their evaluation contract. Also, the external evaluator is likely to have difficulty working sufficiently close and cooperatively with the agency staff to develop adequate evaluative criteria for individual programs or projects. But, the external evaluators are less likely than internal evaluators to become attached to certain evaluation techniques or designs and to tend to ignore utilizing any new designs or techniques.[40]

Evaluators "versus" administrators. Whether internal or external to the agency, project, or program being evaluated, evaluators are likely to differ markedly from administrators in their approach to evaluation research. The major points of contention are listed in Table 2. Even though many administrators and evaluators would have the same stance on these points, there is a

Table 2. Evaluators "Versus" Administrators: Points of Contention

Points of contention	Administrators' interpretation	Evaluators' interpretation
Role evaluative research should play in policy formation:	Minor	Major
Amount of access of evaluators to decision-makers in the policy area:	Little	Much
Practicality of evaluative research:	Little	A great deal
Nature of results sought:	Often ambiguous	Usually clear and concise
Concern about findings embarrassing the organization	Great	Little
Professional orientation:	Toward the organization	Toward the general scientific community
Desire to publish results of the evaluation:	No	Yes
Central goal is:	Service	Research
Interested in solving problems which are:	Immediate	Long-range
Emphasizes:	Uniqueness of their program/organization	Generalizability of programs/ organizations
Statements of goals or objectives of the program-organization must be:	Vague	Explicit
Tends to rely on:	Intuition and conventional wisdom	"Scientific" decision-making
Attitude toward change:	Often static	Often dynamic

[a]Derived from the discussion by Francis G. Caro, ed. Readings in Evaluation Research (New York: Russell Sage Foundation, 1971), pp. 1-34.

tendency for them to vary in the ways indicated in the table. If there is disagreement on even a few of these points in any particular evaluation research project, then the evaluators and the administrators often become adversaries in the evaluation process. If extreme, this can become very debilitative to the evaluation research.

Perhaps the fundamental problem in this respect develops inevitably because evaluation and organization tend to be conflicting concepts.[41] The organizations staffs' highest priority is likely to be maintaining the organization and their programs or projects. Evaluation, by definition, always has the possibility of providing evidence which can be interpreted as threatening to the organization. This alone explains much of the natural resistance to evaluation research or to suggestions of alternative programmatic strategies.

A great number of the problems concerning the relationship between evaluation researchers and bureaucrats could be substantially alleviated if the idea of an "experimenting society" could be adopted. The concept comes from an article by Donald T. Campbell in which he proposes that reforms be conceived as experiments which should be retained if successful, and abolished if unsuccessful.[42] This is contrasted with the static society in which reforms are treated as if they are automatically assumed to be successful. The experimenting society idea would shift the burden of proof to those who implement programs to demonstrate their success, rather than to its critics to demonstrate failure. It seems somewhat ironical for a project to ensure its survival not by proving itself successful, but by avoiding discovery of its shortcomings.

Social scientists and evaluation research. The increasing involvement of social scientists in evaluation research raises particular problems for the research teams involved in evaluation research.[43] Questions often arise

relating to "applied" versus "basic" research, the potential value of social scientists' contributions to evaluation research, the complexity of the evaluation designs they demand, the difficulties of multi-disciplinary research, and other problems.

The applied versus basic research controversy, and the differential values imputed to each, is a real argument to some social scientists, a "straw man" argument to others. In crass moments, one can reduce the arguments on each side to straightforward accusations of the basic research proponents being merely purists and the applied research proponents being merely charlatans. Actually, even though basic differences do exist between the two, much of the major inability of the two to interrelate is a matter of the socialization received in varying graduate programs in the social sciences.

The most accurate position in the controversy is this: both applied and basic researchers can make significant contributions, and the social scientist who can work competently in both types of research is most likely to be successful in evaluation research.[44] Both types of research are interesting, challenging, and have contributions to make generally, and to each other particularly. For those who doubt the worth of any research which is not reportable in a refereed journal, Rossi and Williams' admonition was well taken when they said:

> . . . a good evaluative study may be worth more to society than the total yearly wisdom set down in a prestigious academic journal.[45]

In the same paragraph, these authors note that two of their recent evaluations have culminated in publications in outlets viewed as being prestigious. Hence, their case argues strongly that the basic versus applied research dichotomy is overly simplistic, and, beyond minimal clear definitional differentiation, is, in the main, a "straw man."[46] May and Cohen[47] have carried this interpretation to its logical, and practical, conclusion. In commenting on developments in

the mental health field, they noted:

> 'Basic versus 'applied' research is not the issue. Neither type of research is utilized as much as it should be. In neither case is there a well conceptualized system for integrating research into practice, and the interface between scholarly scientific knowledge and social science and psychological professional practice is at best haphazard, and at worst a moat around an ivory tower. The situation is likely to become even worse as our technology evolves and becomes more complex.[48]

Feedback problems. Agencies generally are not organized nor oriented toward taking evaluation research findings into the decision-making process. Regardless of whether external or internal evaluators are utilized, it is difficult to organize to ensure sufficient bureaucratic dynamism for capitalizing adequately on the findings of the evaluation research.

It is difficult to time the availability of the evaluation research results to the cycle of decision-making which will maximize the utility of the results. Adequate designs and write-ups often get extended in spite of the most careful initial planning. Even though early feedback of preliminary results may seem advantageous to the immediate action program, and even be tempting as a technique for holding the administrator's attention and support, such a step often endangers the entire evaluation procedure if it has been cast in a quasi-experimental or experimental research design.

Funding levels. Adequate evaluation research is expensive. Our society has no qualms about spending billions on evaluation research of weapons systems, but it has had difficulty bringing itself to invest adequate funds to evaluate our systems of human services.

Even short term or one-shot evaluations can be very expensive where surveys and other extensive data collection or analysis techniques need to be employed. The time element, panel techniques, and other techniques add to the expense of long range evaluation research.

Administrators, evaluators, politicians, and the public have to be committed to long range funding guarantees if comprehensive and rigorous evaluation is the desired goal. In the long run they probably have no choice other than making their commitment. The current practice of continual funding for programs with inadequate understanding of the outcomes is likely to continue to lead to policy disappointments of the most severe type. The consequences of this procees is obvious and very unhappy for all parties concerned.

V. DISCUSSION

In differentiating throughout the paper between evaluation and evaluation research we have argued for systematic and rigorous research based on scientific procedures. From this perspective, evaluation research has the chance of being sufficiently scientific to produce generalizable results and, thereby, to lead to theory-building in the area of evaluation as well as the substantive area under consideration. This cumulative process facilitates the development of standard interpretations based on hard evidence which can guide policy-making.

The methodological considerations and evaluation techniques presented are a guide to developing research designs which incorporate at least minimal scientific standards for evaluation research. Of course, all of the canons of scientific inquiry cannot be met in all evaluation efforts but if as many as possible of the basic guidelines are followed, the rigor of the design can be improved to minimize bias.

The political considerations presented are probably a more difficult barrier to evaluation research than are the methodological problems. Political, ethical, and organizational considerations impede evaluation research significantly. These impediments can be ameliorated through a real development of awareness of the problems by all parties concerned. Also, researchers, administrators, politicians, and the public need to be given information about the necessity for evaluation research. We do not conclude, as have some authors, that the organizational problems are so great as to negate the worth of evaluation research.[49] On the other hand, the most serious aspect of this problem

may be the tendency of administrators to oversell the objectives and goals of their programs or projects, and the analagous tendency of evaluation researchers to oversell the payoffs of evaluation research.[50]

It is true that evaluation research cannot be accomplished adequately at times because of these problems; similarly, its findings are often utilized ineffectually--if at all. We believe, however, the difficulty of these practical problems cannot be allowed to deter evaluation research because of its fundamental importance to the development and implementation of adequate services in the areas of human resources.

FOOTNOTES

1. James E. Anderson, Public Policy-Making (New York: Praeger, 1975), p. 5.

2. Ilene N. Bernstein and Howard E. Freeman, Academic and Entrepreneurial Research (New York: Russell Sage Foundation, 1975), p. 1; Joseph S. Wholey, John W. S. Canton, Hugh G. Duffy, James S. Fukumoto, and Leona M. Vogi, Federal Evaluation Policy: Analyzing the Effects of Public Programs (Washington: The Urban Institute, 1970), p. 23.

3. Edward S. Suchman quoted in Francis G. Caro, "Evaluation Research: An Overview," in Francis G. Caro, ed., Readings in Evaluation Research (New York: Russell Sage Foundation, 1971), p. 3.

4. Anderson, Public Policy-Making, pp. 132-134.

5. Bernstein and Freeman, Academic and Entrepreneurial Research, p. 18.

6. Johan Galtung, Theory and Methods of Social Science Research (Oslo: Universitetsforlaget, 1969), pp. 452-453. David G. Leege and Wayne L. Francis, Political Research (New York: Basic Books, 1974), pp. 4-5.

7. Wholey, et al., Federal Evaluation Policy, p. 24.

8. Some of these categories are borrowed in modified form from Wholey, et al., ibid.

9. These categories are modified from those given in Bernstein and Freeman, Academic and Entrepreneurial Research, pp. 18-21.

10. For a more comprehensive discussion of natural vs. analytic systems, see David Easton, A Framework for Political Analysis (Englewood Cliffs, N. J., 1965); A. F. Wessen, "On the Scope and Methodology of Research in Public Health Practice," Social Science and Medicine, 6 (1972), pp. 469-490 presents a useful discussion of this conceptual problem within the discipline of public health. Although on a slightly different topic, Mark G. Field, "The Concept of the 'Health System' at the Macro Sociological Level," Social Science and Medicine, 7 (1973), pp. 763-785 is a provocative conceptualization of the "health system" as a natural system.

11. Easton, A Framework for Political Analysis, pp. 47-48, 59-61.

12. Edward A. Suchman quoted in Francis G. Caro, "Evaluation Research: An Overview,"In Francis G. Caro, ed., Readings in Evaluation Research.

13. Interesting and provocative attempts to measure need as it pertains to medical care have been provided by A. George Gitter and David Mostofsky, "Toward a Social Indicator of Health," Social Science and Medicine, 6 (1972), pp. 205-209; Mark E. Schaefer, "Demand Versus Need for Medical Services in a General Cost-Benefit Setting," American Journal of Public Health, 65 (1975), pp. 293-295; Donald L. Patrick, J. W. Bush, and Milton M. Chen, "Methods for Measuring Levels of Well-Being for a Health Status Index," Health Service Research, (1973), pp. 228-245.

14. Particularly important to evaluation research in health care is the concept of health culture. For an interesting discussion, see Hazel Hitson Weidman and Janice A. Egeland, "A Behavioral Science Perspective in the Comparative Approach to the Delivery of Health Care," Social Science and Medicine, 7 (1973), pp. 848-853.

15. For a discussion of this see William Mendenhall, Introduction to Probability and Statistics (North Scituate, Massachusetts: Duxbury Press, 1975).

16. The requisite level of sophistication of an experimental design will depend on the nature of the program under study, the variation among subjects, and the amount and accuracy of information required. For example, to eliminate the problem of variation among cats noted, the experimenter may construct a randomized block design. In this type of design each cat is given all four treatments. Obviously, this type design can only be implemented if the effects of the drugs are relatively short lasting so that the applications are independent. The advantage of this design is that the four treatments are measured on homogeneous units--that is a single cat receives all four treatments. The size of the sample of cats receiving these four treatments depends on the accuracy desired but it will be smaller than that of the completely randomized design.

    A modification of the randomized block design is the Latin square design. In the former, randomization referred to the order in which the drugs are applied--the four drugs being applied in random order. If there is a change in the cats over time, such as fatigue, and this change generally affects all the cats, the experimenter may try to counteract this trend by means of the Latin square design. Noting that for any cat four drugs must be applied in four different time intervals, the Latin square design requires that for the experiment no two drugs be applied in the same interval. This design requires a sample of four cats, or a multiple, say k, of four cats may be used in which case each drug would be applied k times in each time interval. This design guarantees that no drug is over or underrepresented in any time interval.

    The progression from a completely randomized design to a randomized blocks design to a Latin squares design was made to eliminate specific sources of variation. If the drug treatment wears off over time, the randomized block design can be used to decrease inter-subject variation. Similarly, the Latin squares design can be used to decrease the effects of any variation that occurs systematically over time among all the subjects.

    More sophisticated experimental designs which provide more information can be constructed. For example, the randomized block design can be

replicated--that is each of the four drugs can be applied more than one time to each cat. This design allows the experimenter to determine if there are any interactions between the cats and the drugs. For example, does the blood pressure of one cat increase when drug x is replaced by drug y while the blood pressure of another cat decreases when the same change is made?

17. The systematic change in the experimental units cited in fn. 16, which the Latin Square design is meant to counteract, is one example of such bias.

18. A randomized block design would be possible in this instance only if groups of persons which are considered uniform with respect to relevant factors (for example, medical history) are considered as blocks. The drug and placebo could then be administered to different individuals within the block. Individuals could not be considered as blocks because of the lasting effects of the drug.

19. George W. Torrance, et al., "A Utility Maximization Model for Evaluation of Health Care Programs," Health Services Research, 7 (Summer, 1972), p. 118 ff.

20. Marvin R. Burt, Policy Analysis: Introduction and Applications to Health Programs (Washington: Information Press, 1974).

21. See Torrance, "A Utility Maximization Model for Evaluation of Health Care Programs," for a discussion.

22. Ibid.

23. Ibid., p. 119.

24. A. R. Prest and R. Turvey, "Cost-Benefit Analysis: A Survey," The Economic Journal, 75 (December, 1965), pp. 68-73.

25. Ibid.

26. A public good, such as defense, is a good which, if supplied to one member of a group, is supplied automatically to all members of the group.

27. For example, see: Alice Rivlin, Systematic Thinking for Social Action (Washington: The Brookings Institution, 1971), p. 108 ff; Peter H. Rossi and Walter Williams, ed., Evaluating Social Programs (New York: Seminar Press, 1972); Caro, ed., Readings in Evaluation Research; Bernstein and Freeman, Academic and Entrepreneurial Research.

28. See. B. G. Greenberg, "Evaluation of Social Programs," in Caro, ed., Readings in Evaluation Research, pp. 155-174.

29. Barbara Goltz, et al., "A Built-In Evaluation System in a New Community Mental Health Program," American Journal of Public Health, 63 (August, 1973), p. 702.

30. Wessen, "On the Scope and Methodology of Research in Public Health Service," p. 474.

31. Donald T. Campbell, "Reforms as Experiments," American Psychologist, 24 (April, 1969), reprinted in Caro, Readings in Evaluation Research, pp. 233-254.

32. One of the better known examples of this has been the resistance to the evaluation results of the Head Start program. (See Walter Williams and John W. Evans, "The Politics of Evaluation: The Case of Head Start," The Annals, 385 (September, 1969), pp. 118-132. Another example is the attack on the "Coleman Report," see James S. Coleman, "Reply to Cain and Watts," in Rossi and Williams, eds., Evaluating Social Programs, pp. 97-107.

33. For a review of this problem, see the discussion by Shirley S. Angrist, "Evaluation Research: Possibilities and Limitations," The Journal of Applied Behavioral Science, 11 (#1, 1975), p. 77 passim.

34. Several of the works cited deal with this problem. See Rossi and Williams, eds., Evaluating Social Programs; Angrist, "Evaluation Research: Possibilities and Limitations"; and, Bernstein and Freeman, Academic and Entrepreneurial Research, p. x.

35. For a detailed presentation of the Federal evaluation efforts, see Wholey, et al., Federal Evaluation Policy.

36. In the main, the discussion of this problem follows that of Caro, Readings in Evaluation Research, pp. 17 ff.

37. Caro, ibid.

38. Wholey, et al., Federal Evaluation Policy.

39. Caro, Readings in Evaluation Research, p. 17.

40. For a study illustrating this problem read the article by R. David Mustian and Joel J. See, "Indicators of Mental Health Needs: An Empirical and Pragmatic Evaluation," Journal of Health and Social Behavior, 14 (March, 1973), pp. 23-27.

41. Aaron Wildavsky has developed this idea extensively and is persuasive on this interpretation. See his article, "The Self-Evaluating Organization," Public Administration Review, 32 (November-December, 1972), pp. 509-520.

42. Campbell, "Reforms as Experiments," in Caro, ed., Readings in Evaluation Research, p. 234.

43. In the health evaluation research area these problems are not confined to the U. S. See the discussion of some of these problems cross-nationally in the article by Manfred Pflanz, "Relations Between Social Scientists, Physicians, and Medical Organizations in Health Research," Social Science and Medicine, 9 (1975), pp. 7-13.

44. Although the interpretation seems an overstatement, some commentators have argued that social scientists' basic research offers little to evaluation research. For example, see Walter Williams, Social Policy Research and Analysis (New York: American Elsevier Publishing Co., 1971).

45. Rossi and Williams, eds., Evaluating Social Problems, p. xvii.

46. An interesting side development of this controversy has been the development of a number of scholarly journals for the "applied" area. For example, see Health Politics, Policy Studies Journal, Policy Science, Evaluation, and the forthcoming Journal of Health Politics, Policy, and Law.

47. Philip R. A. May and Jerome Cohen, "Development Operations in Mental Health Delivery Systems: An Urgent Need," American Journal of Public Health, 65 (February, 1975), pp. 156-160.

48. Ibid., p. 156.

49. For an example of this interpretation, see the work of Bruce A. Rocheleau, "The Organizational Context of Evaluation Research," mimeo., a paper prepared for delivery at the 1974 Annual Meeting of the American Political Science Association, Chicago, Illinois, August 29-September 2, 1974.

50. See Rossi and Williams, eds., Evaluating Social Problems, p. 16 ff.

EVALUATION AS AN ADMINISTRATIVE TOOL

Vlado A. Getting, M.D., Dr.P.H.
Chairman and Professor, Department of Allied Health Sciences
Florida Technological University

Emeritus Chairman Department Community Health Services
Emeritus Professor Public Health Administration
School of Public Health, University of Michigan

Presented at
Conference on
Evaluation Research in Health Care
With Special Application to the Florida Division of Health

University of South Florida
June 16, 1975

In 1964, just 11 years ago, Albert V. Hardy[1] said: "In public health administration there will be notable advances in cooperation. The voluntary and official health agencies will have grown closer together, each sharing in a partnership seeking common goals. The health related activities in other governmental agencies will be a united part of the total fabric of public health. The role of the State Board of Health as the leader in community health matters will be freely acknowledged. However, the broad accomplishments for the protection and improvement of thepublic's health will be the result of team endeavor."

Dr. Hardy[2] continues in Millstones and Milestones, Florida's Public Health" as follows: -- there are hopes -- we dream. There will be a wider interest in research and a critical continuing evaluation of the efficacy of all programs". These are the prophetic words of our friend and associate in Chapter X entitled "Looking to the Future." How true they have become! Each of us can testify to the need for cooperation, a partnership and team endeavor for the attainment of common goals. No longer can a health agency, voluntary or governmental, plan, implement or evaluate in isolation.

Everywhere about us we hear the terms:

Accountability
Cost containment
Productivity
Cost-benefit
Evaluation models
Systems analysis
Efficiency
Effectiveness
Appropriateness
Adequacy
Priority and many similar terms. All are geared at determining "is the orange worth the squeeze".

[1]Hardy, Albert V.: Millstones & Milestones, Florida's Public Health from 1889, Florida State Board of Health, Monograph #7, 1964.

[2]Ibid.

"Government at all levels is expanding its contribution to the payment of health services and medical care costs, both in absolute terms and relative to the total investment in these services."[3] For years we have talked about "creeping federalism" into the health industry. By now the federal's continuious increased pace of involvement may be described better as "giant steps". In words known to all: "He who pays the fiddler calls the tune."

With the increasing expenditures of federal funds in health care, there is increased control as to how these funds shall be expended. From cries of utilization review, quality of care and cost containment we are constantly required to justify the expenditures of federal and matching funds in a manner which would:

assure accountability
provide high standards of care
meet acceptability, accessibility and availability expectation of a wide range of beneficiaries. Services must be provided in an effective and efficient manner, be appropriate to the situation and adequate for the size of the problem.

Since all public health administrative services are delivered on a community-wide design, planning, implementation and evaluation of health programs becomes a cooperative and coordinated process for the entire community and not just the administrative health agency. The program is no longer the brain-child of the health officer or some other individual. The program is not the creation and responsibility of a single agency.

"Health planning is a dynamic process, a means to an end, which is concerned with identifying problems, considering alternatives, and making

[3]Thorner, Robert M.; Health Program Evaluation in Relation to Health Programming, NSMHA Health Reports Vol. 68 p. 525-532, June '71.

decisions about future actions, including implementation and evaluation of the plans evolved by the process.[4]

Health planning must become the participative responsibility of all concerned; from the most concerned the recipient of the service to the least concerned those who pay for the service. All in the community must actively and productively participate in each of the three basic phases of programming: planning, implementation and evaluation. Thus the program becomes a community creation -- an accepted responsibility of its conceivers all of whom have a proprietary and paternal interest in seeing that their program, their creation, their brainchild is successful. Other people's babies -- they may be ugly -- but your baby -- that's beaudiful. To be successful a program then should have the attribute of being broadly conceived, by receipients, providers, community leaders, representatives of all groups, segments, subcultures, agencies and relevant community segments.

In all governmental expenditures for health especially now with severe economic constrains the need for accountability and increased productivity with concurrent cuts in resources are being demanded. Equally, united appeals and other voluntary health organizations are faced with increased demands for services and decreased resources. Overutilization, underutilization, duplication, low productivity -- are some fo the no-no words. Thus all health programs, documents, and applications for grant support of projects for the development or conduct of health programs, contain a section on evaluation. Frequently, there is a requirement by grant review groups that the evaluation must measure end results -- either on the health status of persons served or on the state of the environment. Thus to count activities, or the number of

[4] Brown, Anna B. and Getting, Vlado A.; A Manual for Health Planning, Univ. of Michigan Community Health Services Study #6, Univ. of Michigan, Ann Arbor, Michigan, 1969.

services rendered is inadequate. The grant reviewers want to know how the grantee will measure the effect of the proposed program or project on the improvement of the health of the community.

Programming[5] can provide one of two ways of improving the delivery of health resources. One is to increase resources -- a method which is currently not feasable. The second is to "make the best possible use of those (resources) available". The basic purpose of programming is to get the best and most for the dollar expended.

In an attempt to provide the county health officer with assistance in programming, the Florida Division of Health has developed the "County Health Director's Program Guide".[6] Divided into Program Sections the document attempts to present each section under a series of rubrics entitled:

Historical Notes
Philosophy
Legal
Objectives
Programming
Assistance Available from the State
Minimal Acceptable Program
Criteria for Evaluation

However, there are considerable differences among the several sections and there are no definitions or uniformity of proposed methodologies, models or even suggestions of the programming process including the P.I.E. sequence. Hence there can be no uniformity of developments of programs and comparisons among the counties become difficult and at times meaningless.

Time does not permit a discussion of planning or implementation -- each of which is a phase as is evaluation in the total cycle of programming. Our purpose is to consider evaluation not as a terminal activity but as an

[5]Ahumada, Jorge et. al; Health Planning, Problems of Concept and Method; Pan American Health Organization; Sc. Pub. # 11, Washington, D.C. 1967.

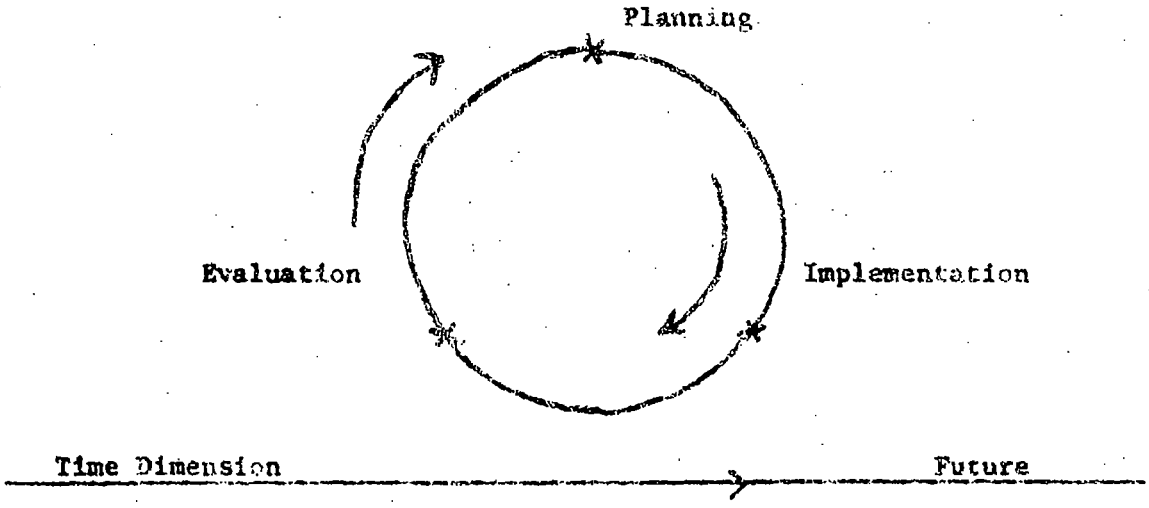
[6]Florida Division of Health: County Health Director's Program Guide, Jacksonville, Florida, Circa 1970. Revised periodically.

ongoing process. Thereby evaluation becomes an administrative tool designed to improve program effectiveness, efficiency, appropriateness and adequacy.

A Conceptual Framework for Programming

Whoever engages in programming should have a firm grasp of the programming cycle realizing that it is a process by which people make decisions about the future. The end, then, is not just to develop a plan, but to produce action towards a rationally calculated goal.

The fundamental elements of the programming cycle in planning, implementation, and evaluation,[7] as shown in Figure 1.



[7]Getting, Vlado A. "Planning, Implementation, and Evaluation of Community Health Services." Canadian Journal of Public Health, Vol. 55, No. 12 (December, 1964), 513-21.

Sometimes it is helpful to subdivide the elements further to illustrate the steps necessary for action. See Figure 2.

Fig. 2.--Health planning:

1. Awareness of Problems and Needs

STEPS FOR ACTION

2. Define Problems and Needs

3. Collection of Data

4. Goals (Long-Range)

5. Determining Priorities

One must always ask--"What are the consequences of doing nothing about the problem?"

* * * *

Changes may be expected at any step due to changes in the problem, in technology, changes in objectives due to additional data and progress in meeting objectives, shifts in action plans, awareness and attitudes of the community, etc.

6. Specific Objectives

7. Choosing Appropriate Activities

8. Allocation of Resources

9. Implement Action Plans

10. Ongoing Evaluation

Time Dimension ——— OBJECTIVES REACHED

Progress Towards Goals

Figure 2--continued

Relationship of Steps to Pertinent Questions
to be Asked at Each Step

At every step questions have to be asked, and decisions have to be made. Briefly, some pertinent questions relative to each step are:

Step 1. Awareness of problems and needs.--Is there a problem or need? Who thinks so?

Step 2. Define problems and needs.--What is the problem or need? Who defines or helps define?

Step 3. Collection of data.--What is the evidence? How big? How small? How severe? How costly? Is more data necessary to substantiate the problem? How much data is necessary before one plans to attack the problem? What is the distribution of the problem? Where is it concentrated?

Step 4. Goals (long-range).--Given what is known, what can be done about the problem? (goals) What are the restraints such as resources, technology, etc.? What is the feasibility--is there knowledge that would assure a successful attack on the problem?

Step 5. Determining priorities.--For each problem: Is there a lack of resources? Is there a lack of technology? What are the consequences of doing nothing about the problem? Given all the problems and the restraints, which ones should be tackled first, next, etc.? What are cost-benefit ratios? Societal and community values?

Step 6. Specific objectives.--For each priority established: What are the objectives? Are they clear? Are they stated in quantifiable terms? What is the expected outcome for each? What is the time limit? Where and in what order?

Step 7. Choosing appropriate activities.--Based on the objectives: What are the alternative ways (methods) to attain each objective? What resources are available for each alternative? What are the consequences for each method? What is the best method (most appropriate activity) to use? Is it realistic, feasible, attainable, and acceptable to clients, to professions, etc.?

In the development of action plans--the ways or methods--consider such questions as Who does what? To whom or for whom? With whom? Where is it to be done? For how long and how much? What will it cost? What are the restraints? How or can they be circumvented? How do you measure the effectiveness of the action? How is coordination of agencies obtained?

Step 8. Allocation of resources.--Are the resources properly allocated? Personnel, facilities, equipment, and supplies?

Step 9. Implement action plans.--And when the action begins, plan to evaluate: Is the plan working? If not, why not? How is the progress to be measured?; and ask as well: Who is doing what? To whom or for whom is it being done? Where is it being done? For how long and how much? What is it costing? Are there any restraints developing? Can they be circumvented? Are the observed results due to the program or to other factors?

Step 10. Ongoing evaluation.--At the end of specified periods of time ask: Is the action plan working? If not, why not? What needs to be revised to reach the objectives? When the progress is measured, are the objectives being reached? In the advent of new data, technology included, do the goals (or objectives) need to be redefined? Is evaluation used as an operational tool to find roadblocks and failures, to make necessary corrections?

--And the cycle begins again.

There is a logical flow from element to element, and decisions have to be made at every step of the way. Steps may emerge and flow into one another--often in actual practice less clear-cut than diagrammed. Decisions made at any step may influence both past as well as future decisions--thus there is continual interaction among the elements in the whole process. This means that evaluation is really not a separate end point, but is an ongoing operational process related to each and every element.

Programming is a decision-making process, selecting from various alternatives, culminating in action, and operating within the framework of the values, and attitudes of the community. The process begins with a feeling of dissatisfaction with things as they are, definition of problems and needs, collecting necessary data, setting long-range goals, determining priorities, setting specific objectives, choosing appropriate activities, allocating resources, and culminating in action, implementation, and evaluation--the measurement of progress.

Goals, used in this context, refer to broad statements regarding a desired state of affairs which extend over a long time range. Given a number of goals pertaining to different problems, decisions must be made about which goals are most urgent; that is, priority setting. If resources--money, personnel, time, etc.--were unlimited, there would be no need to set priorities. Sometimes, a lack of technology is a restraint affecting priority setting.

Objectives are then determined according to priority choices. The term "objectives" refers to very specific quantifiable statement which describe what is expected to result from action.

"Only in the last decade or two has serious attention been given to evaluation of the various forms of organization of health services".[8]

[8] Roemer, Milton I; Evaluation of Health Service Programs and Levels of Measurement, NSMHA Health Reports, Vol. 86 #9 p. 839-848, Sept. 1971.

The evaluation of health programs is complicated by an admixture of variable. Hence there has been a great deal of confusion in deciding what should be evaluated, let alone how to go about doing it. The term evaluation has a wide disparity of meanings. In this presentation it is used in a framework for relative values of various systems, subsystems, patterns of program operations. Programs may be evaluated at various levels. Roemer[9] describes six levels of results, as is illustrated in the diagram. Evaluation may be used for one of several purposes. Evaluation can be considered formally from the point of view of systems control theory.[10] "This implies that evaluation must be an ongoing process within the system, linked to decision-making levels, and employing both feedback and feedforward of information". Thus the evaluation process must be developed along with the planning of the program and steps be provided to accumulate essential data, analize and interpret it, provide periodic reporting to decision makers so that it can be utilized to determine what changes if any need to be made to assure the attainment of the predetermined quantified objectives.

Such ongoing evaluation may result in redefinition of objectives, reassement of the problem, identification of external or independent variables, reconsideration of alternatives, reassignment of resources, alteration of selected activites, reassignment of responsibility and authority, changes in data collection, priorities or a host of other factors that may improve the program.

The methodology of ongoing evaluation is prospective-- the evaluator needs to determine the data he will need, how it is to be collected, what is to be the

[9] Ibid.

[10] Zemach, Rita. Program Evaluation and System Control; A.J.P.H. Vol. 63 p. 607-9 July '73.

method and content of his periodic reporting to the program's decision makers. This is not a retrospective evaluation study; such end-of-program evaluations often are incomplete because essential records were not identified and therefore not accumulated. Such a study is therefore usually limited in scope and designed at most to give some tailend judgement.

Prospective evaluation on the other hand is designed to be an operational or administrative tool which is developed before the program becomes operational. It is based on the accumulation of data and is therefore objective or factual and not subject to subjective analysis. The purpose and intent of the evaluation is defined in advance. It forces better planning and better implementation if the findings are applied by the decision makers to modify the program. Prospective evaluation is a time consuming and costly process that requires expertise. It is usually limited to certain specified phases of selected programs. The evaluator must be knowledgeable in in administrative systems control as well as in public health and must have access to the decision makers.

Before ongoing evaluation can become an operational tool several questions need to be answered:

1. What are the real objectives of the program? Economic? Political? Educational? Humanitarian? Prestige of individuals? Unless evaluation deals with the real objectives, its findings will be of little concern to decision-makers. Honesty in stating the purposes of a program may at times be painful. But if a program is established to "do something visible in the inner city" then it deserves to be evaluated in terms of its visibility.

2. To whom must evaluation be addressed? Administrators? Board of Directors? Professionals? Legislators? Consumers? A federal agency? There may be many individuals and groups whose judgments have bearing on the fate of the program, and each may be seeking different information. Thus evaluation must be multidirectional, responding to many, perhaps conflicting, interests.

3. What types of evaluation criteria will be used? Services rendered? Cost reduction? Approval of professionals? Satisfaction of recipients? Suffering alleviated? Political payoff?

4. Is there competition for resources among various objectives? How will allocation be determined and according to which criteria?

5. What are the time factors? How soon must progress (or its lack) be demonstrated, and to whom? How ofter will reports be expected, and what types of observations will relate to this time span?

6. What methods will be used to generate information? Review of records? Observation of performance? Interviews?

7. Is the desired-output or reference fixed, or is it changing over time? Are the new objectives to be met? Are broader aims superseding the internal program objectives?

8. Is progress of individuals measurable in relation to program objectives, or aggregate progress, or both? And how?

9. How much of the time and effort of staff, and how much of the total budget, can be allocated to support of evaluation activities?

10. Must the total program itself be justified, or is this a new organizational or technological configuration for an accepted program?

11. What explicit channels are needed for feedback of information to those who require it? [11]

What is the purpose of the ongoing evaluation? Are we merely fulfilling the needs of reporting to the State or Federal government and therefore evaluation is a bore and to the local health officer a necessary nuisance. On the other hand, ongoing evaluation designed as an operational tool can serve not only the local health officer, but the program director and the entire program staff. In this context of an operational tool, prospectively designed evaluation properly applied can improve the implementation of any program, system or subsystem.

[11] Ibid.

SUMMARY

Evaluation is a part of the programming process of planning, implementation and evaluation. Prospectively designed evaluation methodology can become an effective systems-control device designed to improve productivity. If program productivities are to be compared, there should be standardizations of the evaluation methodology.

Delivered by Dr. Hyatt

Evaluation of Health Services

As of July 1, 1975, a new era in the delivery of health services is being initiated in Florida. Foremost in our consideration of evaluation of these services must be appropriateness in the legislative directions for the reorganization of HRS.

The dominant purpose of this reorganization is the integration of the delivery of all health, social and rehabilitative services. Though attention is directed to "clients" in need of assistance, there is repeated reference in the statement of purpose, to prevention. Included is a mandate to prevent dependency, prevent abuse of children, preserve families, prevent inappropriate institutional care, prevent communicable diseases and other physical and mental disease to the maximum degree possible, promote health, disseminate health information for the benefit of all residents and visitors in Florida and assure high quality of health services fully accessible to all citizens.

The unique feature in the legislative plan is the organization of services on a district basis with critical attention to program planning and the maintenance of uniformly high standards. The design for the operation of this program by more than 30,000 employees is specified in the legislation. Its approaches are new and the dominant question in the minds of legislators is how well will it work and of employees will it work. To answer such questions, Program Evaluation is specifically required. This is an activity of such importance that responsibility for it is placed in the Secretary's office, with provision that he may assign duties to any appropriate unit within the department. Moveover, it is provided also that whenever possible, DHRS shall contract for services if such services are more cost efficient than those provided by the Department. These

are considerations to be kept in mind in present discussions and planning.

Evaluation as visualized by the legislators is spelled out in the reorganization legislation as follows: (Quote Section 5, Program Evaluation ---- to the end of quote).

It is doubtful whether the magnitude of the tasks specified in this legislation were visualized by those who so clearly specified what would be desirable. Faced with a task which has never been adequately accomplished, there will be a temptation to direct attention to what can be counted and the cost of which can be estimated. Hopefully, this narrow cost accounting approach to satisfy a legislative requirement will be resisted and that there will be a determined effort to develop a comprehensive program evaluation which will be a major guide to increasing excellence.

Whatever the evaluation method employed, the point of departure is a knowledge of the purposes and objectives of reorganized HRS. As previously indicated, a major purpose is the integration of the delivery of all health, social and rehabilitative services. In the health field alone, present programs in public health, mental health, childrens medical services and the medical and health aspects of other Divisions are operated independently. Evaluation will need to indicate the effectiveness of these services when combined in clinics and the community. Can a nurse or other worker in a home visit satisfy multiple purposes in the interest of efficiency and client satisfaction. The effectiveness of this approach will not be measured by numbers and costs. The district organization is a second major modification, and the effectiveness of this needs to be assessed by objective and unbiased evaluation. Innovative also is the separation of the responsibilities for program operations from program planning and development. Is this a practive to be recommended? As a guide to indicated modifications and as an objective record

of accomplishments, these changes need to be critically evaluated. This is but a beginning of what is expected since a comprehensive program evaluation is to be established for all major programs. What recommendations can we offer?

Little guidance can be provided by past experience. In public health the importance of program evaluation has been endorsed in words though not often formally recognized in practice. Administrators have assessed the quality of programs with a generally dependable feel, bolstered by public reaction. Evaluation schedules have been developed by the American Public Health Association and for a time were utilized by the exceptional rather than the average or weak departments. More visible have been the special program reviews and evaluations by individual consultants or a committee of consultants. A personal experience sharing in an evaluation of the Public Health Program in Los Angeles County, leaves me with a favorable impression of the effectiveness of this approach.

Looking to the future, the establishment of measurable objectives and performance criteria for all health programs will be a new and difficult planning requirement. The accounting of costs of services and determining the relative costs and benefits of differing procedures will involve complex accounting. Particularly difficult will be evaluation of the effectiveness of preventive programs as this calls for an estimation of what does not occur. How shall we determine for example the value of programs to prevent child abuse? Equally difficult, how shall we measure the relative values of activities to promote the maintenance and improvement of health and mental health?

In weighing possible procedures for the evaluation of health programs, there are some guiding considerations. The history of public health emphasizes the effectivene and the economy of prevention. The Florida State Board of Health was finally established as an aftermath of a disastrous epidemic of yellow fever in Jacksonville. As a result of subsequent research, the means of spread and methods for prevention were clarified and yellow fever was promptly eradicated from our state and country.

Through prevention, dread cholera and small pox became historical memories. Diphtheria, poliomyelitis, measles, whooping cough, pertussis, and other infections were brought under control as a result of research and preventive action. Acute diarrheal diseases of children and tuberculosis, both were the first cause of death at differing times in earlier years. Both have yielded to general or very specific preventive programs. Prevention has paid off admirably in the control of communicable diseases. Current major probelms also may be alleviated through prevention. It is accepted that prevention is the hopeful approach to our present number one killer, the cardiovascular diseases. The progressive and frightening upward trend of lung cancer would be eventually reversed by a marked reduction in cigarette smoking. The care of the alcoholic alleviates an immediate problem but prevention is needed to solve the medical and social problems of alcoholism. Prevention is the promising approach to the control of chronic diseases. But changes come slowly. It must not be supplanted by care of clients with benefits which are immediately apparent. In evaluation, it will need to be recognized that assessment of the benefits of prevention should be designed to measure changes over a period of years, rather than during single years.

Even though the legislation identified Program Evaluation as a separate activity for which the Secretary shall have responsibility, it seems clear that coordinated action will be indicated. A design for program evaluation will need to be a part of program planning and development. In this process, program objectives and specific goals as well as the performance criteria, quantifiable wherever practicable, will need to be specified in carefully prepared documents. Reviewing and monitoring district level program operations, assuring compliance with program standards, and assuring uniform program quality are specified activities of program planners; all involved program evaluation by these program authorities.

While operations are legislatively separated from evaluation, it seems apparent that the district program managers of health services and the program supervisors will be aware of problem areas as well as the effective operations. This is a useful evaluation. Inevitably, operations personnel will be participants in any program evaluation involving field studies.

It is specified further that when possible, the departments information system shall provide the basic data for program evaluation studies. Those planning for the collection of data and for the analysis and use of statistical information will be contributing to program evaluation.

While this wide participation in program evaluation may be developed, the success of this activity will be dependent on a qualified staff having responsibility for the total evaluation activities. They should provide leadership in obtaining statements of program objectives so expressed that measures of attainment can be obtained. They will provide the expertise for the development of evaluation schedules, make available trained interviewers, prepare and distribute the evaluation reports and aid as requested in the use of evaluation findings for increasing the effectiveness of programs.

This evaluation staff and the evaluation activities may be obtained either by developing an expanded evaluation unit in the Department, or by the purchase of these services, or presumably by a combination of these provided always that there is an advantage in cost and effectiveness.

I see benefits in contracting for some or all evaluation services with an institution such as the College of Social and Behavioral Sciences of the University of South Florida for the following reasons:

Dependable evaluation demands an unbiased approach. This would be better assured by personnel outside HRS than by a staff unit responsible directly to the

Secretary. The latter has the potential weakness of any self evaluation. It would be hard for a departmental staff to avoid emphasizing the favorable. An unbiased objectivity could be expected from a University group reporting to the legislature as well as to HRS.

The focus of attention in the legislative mandate for evaluation is on cost effectiveness. An intra-agency unit could scarcely avoid giving major attention to this. An independent university group in contrast could direct equal or greater attention to the actual effectiveness of programs in satisfying human needs.

Progress in public health in the past has been attained by doing for people, such as insect control, vaccinations, sewage disposal and the purification of water. Major health problems today are related to human behavior as habits in eating, smoking, drinking, speeding, sex relations and in stressful living. Evaluation of control of health and social problems will require an assessment of the effectiveness of programs designed to modify behavior. Moreover, the follow-through of evaluation is program improvement. University based behavioral scientists should be particularly qualified to aid in these problems involving human behavior.

Evaluation needs the service of senior professionals in differing fields. Florida has a largely untapped resource in experienced, able and vigorous retirees. Rarely do they seek full time responsibilities, but many would welcome part time activities which would keep them in touch with life time and continuing interests. I believe there are talented persons available, including retirees from Departments of HRS who could contribute substantially to a dependable evaluation of HRS programs. Tampa is favorably situated geographically to find and use such persons.

Though designed as a service activity, health program evaluation should be approached as program research. The evaluation procedure itself must be the

subject of exacting evaluation. This approach would be favored by university leadership.

Program evaluation must be done in close association with program operations; these will be organized by districts. Tampa is within easy reach of multiple favorable field situations for the development of evaluation procedures, for their testing in pilot trials, and for their use in ongoing District programs.

These comments are personal reactions, not an expression of administrative interests. It is my impression that now, before decisions have been made is the time for proposals. If this or any other institution after careful consideration believes they can provide better program evaluation at less cost than a HRS staff, their interest should be made known. The outcome of the discussions here could lead to specific recommendations and productive action.

Quote referred to on page two:

Section 5. PROGRAM EVALUATION.—A comprehensive program evaluation system shall be established which shall encompass all major programs of the Department of Health and Rehabilitative Services. The department shall establish measurable program objectives and performance criteria for each program it operates. Such system of evaluation shall require all programs to develop identifiable goals which are quantifiable wherever practicable and to estimate the cost of attaining such goals in advance. Studies of the relative cost and effectiveness of departmental and alternative programs shall be conducted. The department shall develop a program evaluation schedule and shall evaluate at least 20 percent of its programs annually. The department shall submit these evaluation schedules and reports to the secretary, who within thirty days of receipt shall transmit copies of the schedules and reports to the appropriate substantive committees of both houses of the legislature for review. When possible, the department's management information system shall provide the basic information for program evaluation studies.

THE FUTURE ROLE OF GOVERNMENT IN HEALTH

and

TRAINING HEALTH MANPOWER FOR THIS ROLE

Report to Dean Myron Wegman

Prepared by Vlado A. Getting, M.D., Dr.P.H.,
Chairman, Department of Community Health Services
Professor of Public Health Administration,
as an off-campus assignment

Presented February, 1974
University of Michigan School of Public Health

THE FUTURE ROLE OF GOVERNMENT IN HEALTH
and
TRAINING HEALTH MANPOWER FOR THIS ROLE

Table of Contents

INTRODUCTION

What are the functions of state and local health agencies likely to be in 1980? What training should a school of public health give its students to prepare them for careers in these agencies so as to assure them the ability to meet the challenges of today as well as those of the 1980's?

Forecasting at best is a tricky business. However, in such a rapidly changing area as the delivery of health care services - including personal, community, and environmental health services - only a fool would attempt to predict the future. Unfortunately, much of the predicting and forecasting which has already been done in this area appears to have been done without taking into consideration the probable shape and climate of society in the future.

There can be no question that the evolution of the health field will be dependent upon a whole host of changes that will take place; social, economic, political, scientific, technological, as well as the inputs from the health industry itself. To anticipated trends, there will be countertrends. It can be expected reasonably that there will be some pendulum-like swinging backward and forward at a different rate and time schedule not only among specific trends and countertrends but that variations will occur geographically.

This forecast of probable future roles of state and local health agencies is further complicated by such recent events as the Middle-East War, the energy crisis, increased cost of living, inflation, and other consequences as reduced productivity and increased unemployment. As national resources become diverted into new areas of concern, such as making the country independent in relation to its needs for energy by the early 1980's, it becomes more obvious that even a crystal ball would not be adequate.

Thus, to project future developments on a straight line based on past performance would be a serious error. The Commission on Education for Health Administration is taking into consideration "Future Trends in Health Care Delivery: A Forecast".[1] Before the final publication of such a paper it undoubtedly will be revised - perhaps several times - and within a year thereafter (1975) it will need revision again.

In a study, with limitations such as are imposed on a single individual geared to a specific time schedule, it is impossible to document opinions, attitudes, or data in this report. I have, therefore, come to the conclusion that the best policy to follow in the writing of the report is not to be all-inclusive but face the difficult job of writing a relatively short document and omit therefrom much of the background material.

This report was requested by the Dean as an off-campus assignment of the writer and it was carried out primarily during the latter half of 1973, although data and work carried out prior to the assignment has been utilized in its completion.

The methodologies employed in developing the study report, are as follows:

1. Review of mass media or popular publications

2. Reading of professional periodicals, books, reports, documents, speeches, and other printed materials

3. Interviews and discussions with individuals in state and local health agencies and faculties of institutions of higher learning

4. Data collected, analyzed, and interpreted in the Illinois Community Health Services Study published by the writer in 1971.[2]

5. Review of my notes and report of a Sabbatical Study carried out from September 1, 1971, through February, 1972.[3] "The purpose of this study was to improve teaching and research in health service administration, to suggest changes, and accumulate materials for other health administration faculty... The methodology used...was a series of study visits to selected areas where innovative health programs could be studied and where schools of public health and/or other institutions of higher learning had developed noteworthy teaching and research in health services administration."

6. A questionnaire (Appendix A) in 1972 which was sent to directors of local health services in state health departments relative to the employment and the training of non-physician public health administrators by state and local health departments.

7. A letter of inquiry (Appendix B) in 1973 which was sent to selected state human resources agency directors, state health officers, state health department chiefs of local health services, selected local health officers and to selected faculties in schools of public health.

The writer wishes to thank Mrs. J. A. Pence for her assistance in reviewing the original document, in sending out the questionnaire and letter of inquiry, in forwarding the replies to the writer, and, finally, for typing and reproducing the report. Simultaneously, I wish to thank Dean Wegman for making it possible to carry out this study. Hopefully, the reader will accept my apology for the style and format of the report. If a few ideas are accepted and incorporated into the changing field of health care service administration teaching, it would be, to me, a real pay-off.

II. The Changing American Scene

"Something, clearly, is stirring. In part we are witnessing changes in personal values that are seen and felt not only in the United States but around the world. In part, we are experiencing the latest chapter in the continuing story of the quest for fulfilling American goals and aspirations: a fair and equitable society; an opportunity for each citizen to participate in the forces that affect his life; a confirmation that the democratic process does, indeed, work for all."[4]

The American scene is constantly in a process of change. Trends and countertrends are taking place simultaneously, at raring rates and speeds and differing from place to place. There is no single trend that is uniform throughout the country. Progress is faster in some places than others, regression takes affect in other places. Various trends and countertrends proceed differently in the same place. On the whole there are certain directions which seem to dominate or prevail.

Changes are taking place in social, economic, political, scientific, technical areas as well as in imputs from the health industry itself. Social changes are related to mobility of families, the increasing dominance of the nuclear family, the changing role of women, increasing population inspite of a lower birth rate and expanded family planning resources, an aging population, a more affluent society, concerns with national critical issues as confidence in government, world peace, the energy crisis, inflation, increased cost of living, decrease mobility due to recent and future transportation limitations, continuing suburbanization and in some metropolitan areas a beginning return to the core city, civil rights progress, desegregation and the progress of all minorities.

Economic changes are reflected in the accelerated rate of continuing inflation, the real increase in gross national product, the dollar value, restricted productivity and increased unemployment due to the energy crisis, increase in minimum wages, increases in family incomes, the need of generating 1.7 million jobs a year through 1980 merely to take care of the new entrants into the labor market and the unemployed. The high cost of living, the unavailability of adequate housing, the increasing portion of the gross national product expended for health, concern for cost-containment, the increasing involvement of government in health are all trends which will affect the evolution of health programs.

Political changes are beclouded by the Watergate and other affairs which have reduced the public's confidence in government officials and government itself. The 1974 election of governors, senators, congressmen and other political leaders will no doubt give some indication of the balance of "liberalism" vs. "conservatism," and may, to some degree, foreshadow the more important 1976 elections which include the President and Vice President. The assumption of the courts of executive or administrative roles and the ongoing battle over authority between the legislative and administrative branches of government are additional factors to affect changes in the health scene. The passage of new legislation and new administrative rules and regulations will set the direction of any health development in the over-all health care or health insurance area.

Regulations of HEW agencies, Public Health Service, Federal Drug Administration, Social Security Administration, etc. will cause changes as these become effective. Other developments include computerized data systems on individuals and communities, reorganization of government agencies, changes in enforcement policies, and alterations in related programs such as funding and direction of education, welfare, agriculture, defense, social security, housing, transportation, etc.

Recent administration economy measures have had a severe and damaging effect on some health programs, especially research and health manpower training. Several court cases have resulted in releasing impounded funds for service, research, training. While the 1974 budget for HEW has not been passed, there are indications that a compromise will be developed between the Administration and Congress. However, because of the late passage of the budget, there probably will be no increased expenditures but more likely continued decreases. Efforts are being made to continue Federal support of many of these programs including the support of schools of public health and traineeships. If such support over the next 5 years is not forthcoming, there will have to be severe cutbacks in schools of public health and some schools may be forced to close. Lack of student support will reduce the number of students not only in schools of public health but other institutions offering training in health services administration, environmental health services, medical, dental, osteopathic, nursing, and allied health at a time when increased personnel will be needed to implement a national health insurance program.

There is evidence that an opinion is gradually taking hold - that more money for health is not the answer. Ample evidence has been accumulated that other countries (Sweden, Great Britain, etc.) are able to provide much more comprehensive health services than does the United States at a much lower percentage of the gross national product (or total per capita expenditure). There is a developing trend in proposed legislation to assure the evolution of a national health policy and a system of delivery that will correct the maldistribution of health resources and inequities of the present polysystem. Pending legislation may expand exploration of additional HMO's (authorizing $240 million) over a five-year period in a "series of additions to our existing pluralistic system." About twenty states have adopted legislation authorizing HMO's, more if not all will do likewise in the next few years. Some concern has been expressed that HMO's may not wish to render health maintainance or preventive services or even delay such elective surgery as cataract operations if their organization is in financial trouble. Experimentation will have to be quite broad so as to assure feasibility. It is not unlikely that a regional authority may be developed to promote, operate and/or supervise HMO's and other local systems that may evolve to deliver personal health services. Such regional organizations may be an existing local or regional health department or human resource agency, a comprehensive health planning B agency, a new government agency, or an extra government unit.

Other legislation and court decisions concerning "informed consent," drug abuse, family planning, abortions, definitions of an adult, drug and medical instruments, restriction of drugs, approval of drugs, development of a drug formulary, automobile safety and pollution, mental health, environmental quality, and a host of others will no doubt bring about changes in health programs.

Scientific and technical developments include social and the physical sciences. From community organization and consumer involvement modalities to surgical, medical, radiological, and other clinical discoveries and the application of new knowledge, skills, and attitudes into the health delivery system. Based on experiences of the past, it may be assumed that these changes will be further accelerated. Further specialization and fractionating among new types of health and allied health professionals, the inclusion of management and system organization type expertise as well as increase in social scientists may be anticipated. The evolution of methodology of motivating people to seek health maintenance and preventive services and of training the health professions to welcome "well persons" as clients will gradually evolve if current trends at cost containment of medical care costs can be taken at face value.

The health industry is a polyglot of professionals, semi-professionals, skilled and unskilled workers. Unionization of hospital workers, most recently interns and residents, will continue to increase costs. There is a dichotomy of views and opinions between physicians and hospitals as expressed by their professional organizations. Some medical groups are concerned about the possible "institutionalization of physician services within hospitals. Great concern has also been expressed relative to the PSRO legislation pending in Congress. A North Carolina hospital has won a case in the state supreme court, overturning a decision of the state that refused to issue a certificate of need for building a new hospital. Differences of opinion exist as to the shortage of physicians forecast for the 1980's. On the whole the medical profession - especially organized medicine is conservative - even ultra-conservative; however, there are indications that attitudes are becoming more liberal. In fact, many physicians, dentists, and osteopaths are developing an attitude of acceptance of the increasing role of government. Some practitioners are resigned to accept what may come, some are taking steps to help formulate policy; programs and their implementation (albeit, this is a mixed pot of porridge). Still others are active in promoting the evolution of programs. Gradually it is being accepted that the consumer, the public, the patient, government, public officials, labor commerce, industry, health industry employees - all are to have a voice in the evolution of health programs.

The rapid evolution of allied health personnel to participate in the delivery of health care indicates that the physician will not be able to continue solo practice - perhaps even in remote rural areas. Regardless of whether HMO's, Health Services Areas, a combination of these or some other patterns of delivery are developed, there will be a need for increasing numbers of allied health personnel.

A continuing resisitance to government "regimentation" or directives can be anticipated. However, if the maldistribution of practitioners and facilities is to be corrected, ways to motivate voluntary or at least temporary obligated geographical relocation must be evolved. Thus, it can be expected that financial assistance to students (health professionals as well as allied health professionals) and financial incentives will attract practitioners as well as health delivery systems to geographical areas that are lacking.

In the final analysis, the determining factor that will decide the health industry's attitude and behavior (mixed as it will continue to be) is the nature of future national health legislation. Some type of national health insurance may be enacted within three to five years, and the actual implementation will probably start 2-3 years thereafter but take several years to be fully implemented. In other words, as national health insurance is evolved, the health industry will gradually adapt to make the required changes.

There is a trend in government to remove physicians from top level policy decision-making and administration and to replace them with political responsive appointees. This reinforces a long-standing position of organized medicine - which was that health departments should not administer health care programs. There has gradually evolved a countertrend, namely, that if government is to become responsible for delivery of personal health care, then physicians in government should be responsible for directing these programs. Thus, there is one trend to remove physicians as administrators and policy makers - now belatedly countered by the medical profession itself. There is a possibility that physicians in governmental health agencies may be relegated to becoming technicians and putting the administration and policy formation into lay hands. Hopefully, these laymen will be trained in decision-making, administration, democratic leadership modalities and community organization.

Because of the rapidly increasing number of court suits against physicians and the award of some very large amounts to patients who won their cases, there has occurred a phenomenal rise in medical and surgical malpractice insurance premiums. Simultaneously, physicians and hospitals have developed a concept of defensive medical practice. Both these factors have increased the cost of medical care. All indications are that such increases will continue. Some medical societies are exploring arbitration as a way of settling malpractice claims without going to court. Although such procedures may effect the size of malpractice insurance premiums, it is unlikely that steps taken to assure an adequate defense (defensive medicine) in case of a suit will be reduced. The trend is to carry out any procedure which may give the attending physician a better defense if he is sued and to not carry out such procedures (as elective surgery) where risk may be high. Thus, defensive medicine costs may continue to increase and the quality of care may continue to suffer even more.

III. The Role of Government in Health

A. The Basic Premises for A Role of Government in Health

"It is a basic assumption of this study that good health is not only a desirable asset in itself, but that it is an essential stepping stone to even more desirable goals; namely, a happy, contented and self-sustaining life. It is assumed further, that without good health in a child, growth and development - physical, social and emotional - may be severely curtailed, leading to poorer education, a lowered earning capacity and increasing dependency and disability. In an adult, lack of good health not only interferes with a person's ability to earn, but may interfer with the earning abilities of other members of his household. Poor health interferes with a person's role in his family and in the community. A person in poor health requires utilization of health services at a higher rate and at an increased cost. He may cease to be a taxpayer and become a public charge."[2]

Further it is assumed that government has not only a direct interest in the health of its people but, with an increasing degree, a responsibility to assist families and communities in providing or seeing to it that there is a provision of a full complement of modern health services. Thus, government, federal, state, regional and local, is supporting to an increasing degree the provision of community health services - both personal and environmental - that are designed to:

1) Assure the maintenance of good health
2) Motivate individuals, organizations and the community to seek good health
3) Prevent and/or control unnecessary sickness and injury
4) Provide primary health services so as to arrest or delay the progress of disease and the onset of complications and disabilities and premature death
5) Furnish comprehensive health services to special groups or persons with special health problems such as:
   a. Infants, expectant parents, mothers and family planning services
   b. The physically and mentally disadvantaged
   c. Oldsters
   d. The chronically ill and homebound patients
   e. Persons at risk to or having:
      1. tuberculosis
      2. venereal disease
      3. communicable disease
      4. drug addiction
      5. alcoholism
      6. kidney disease
      7. cancer
      8. cardiovascular disease
      9. other diseases as lead poisoning, certain anemias, malnutrition
   f. The socio-economically disadvantaged
   g. The migrant worker and his family
   h. Other special population groups

6. Provide rehabilitative services for both children and adults
7. Administration of health care facilities for special, long-term and acute illness
8. Provide supervision, quality control and/or administer public health care programs
9. Provide laboratory support services for personal and environmental health programs
10. Deliver environmental health services to eliminate or reduce health hazards caused by:
    a. Insanitary food, milk, meat, fish and poultry handling
    b. Inadequate housing
    c. Exposure to chemicals and radiation
    d. Unsafe environmental conditions causing accidents
    e. Water, air and noise pollution
    f. insanitary conditions existing in recreation areas, swimming pools, public places, public buildings, schools, etc.
    g. improper sewage and solid waste disposal

It is further assumed that if dollars for community health services* bring about prosperity and happiness; and, at the same time, bring about benefits in reducing costs of medical care that community health service dollars are a real investment - an expenditure with a pay-off.

Based on these assumptions, it is, therefore, concluded that the provision of health services is a well recognized function of government and that this function should be:

1. A partnership with voluntary agencies and private enterprise
2. Available, accessible and acceptable to all people
3. Be adequately financed through a combination of voluntary and private funds - an investment in community health services which give people their best chance for a healthy contented life and which will enable people to flourish not merely exist in a salubrious, health-sustaining environment.

B. The Role of the Federal Government

Since its beginning, the federal government has taken an interest in the health of its people. Provision was made for medical care of our troops even before we became a nation. One of our first official health acts as a nation was the extension of medical care to a special segment of the civilian population - the merchant mariners (the precusor of the P.H.S. now in H.E.W.). However, it was not until after the Civil War that prevention became a concern of the federal government.

*The term community health services as used in this document is defined in Appendix C. It includes all health services except for the privileged relations of a private patient and his doctor.

With the passage of the Social Security Act in 1935, the role of the federal government in the provision of health services was changed to that of a true partnership with state and local governments. The past four decades have seen an avalanche of health legislation and activities which have gradually but consistantly increased the role of the federal government as the senior partner in the provision of personal and environmental health services. "He who pays the fiddler - calls the tune" is now applicable to the health arena. With increasing funding comes increasing rules and regulations; standards and requirements have been set; programs and budgets have to be submitted; reports made periodically and funds and program performance audited. By far the greatest expenditures for health in the United States are made by the federal government and are increasing each year.

The federal government not only conducts its own health delivery programs, but has made funds available to state, regional and local health programs as well as some of the voluntary health agencies. It is the largest source of fees for hospitals and pays directly or indirectly an increasing percentage of fees and capitations to physicians and other vendors. Grants, subsidies and contracts have assisted service, teaching and research programs. Although economy measures for the last two years have reduced some health programs, the overall expenditures of the federal government are increasing each year and will continue to do so as the programs are expanded because of an increasingly large number of beneficiaries as well as larger benefits.

While the direction of health expenditures for health may change it is anticipated that there will be an overall increase. Decreases may occur in training health manpower, in basic research, in the RMP (Regional Medical Program), in capital outlay for health facility construction and renewal (Hill-Burton Program) and in most categorical public health programs. Increases can be anticipated in Medicare and Medicaid expenditures, in general health funds, in HMO (Health Maintenance Organization) funding, in research in health delivery systems and health service administration, emergency medical service programs, substance abuse and perhaps venereal diseases and family planning.

It is further anticipated that some form of "national health insurance" will be enacted in the next 1-5 years. This may bring about a gradual expansion of health services - hopefully with some attempt at correcting existing faults as maldistribution of health manpower and resources, and the evolution of a national health policy and system that will permit innovative and experimental efforts.

Most recently the government has become concerned with the evaluation of activities and programs both in personal and environmental health programs. Interest is expressed in objective planning of specified goals, the listing of activities designed to attain the quantified objectives and in methodology to measure effectiveness and efficiency. Increasing costs and rising taxes have brought about efforts to develop cost-containment especially in the delivery of personal health care services. This effort has led to the idea that health maintenance and prevention may be the key to reducing the demands for expensive medical care at a later date. These trends in better planning, evaluation and in cost-containment are most likely to continue.

C. The Role of State Governments in Health

State governments entered the health arena rather late. It was not until after Johann Peter Frank's six-volume work on public health, "A Complete System of Medical Policy" was published in 1819, Edwin Chadwick's "Report on an Inquiry into the Sanitary Conditions of the Labouring Population of Great Britain" in 1842, and Lemuel Shattuck's,"Sanitary Survey of Massachusetts" in 1850 that state governments in the United States became active in the provision of public health programs. Primarily because of the premature death of persons in the prime of life, the first state health department was established in Massachusetts in 1869, with the direct support of organized medicine. Within a few years, each state had an independent health department or combined health with such activities as lunacy and charity.

Created before the knowledge of the existence of bacteria, they were primarily concerned with the then prevalent concepts of the causation of disease, filth and a lack of knowledge of the laws of hygiene. From these humble beginnings in environmental and educational efforts, they soon expanded into the control of communicable diseases and the collection, analysis, interpretation and dissemination of health data. When the etiology of infectious diseases was firmly established and epidemiological methods and skills evolved, they entered the field of primary prevention and "isolation and quarantine" were supplemented by immunizations. Programs of secondary prevention were also increased, especially in tuberculosis and mental health and, during World War II, in venereal diseases. However, only a few states provided direct personal health services of a therapeutic nature to an individual except in those special instances where there was thought to be a danger to the public health.

There were some notable exceptions, such as Massachusetts and New York. These states operated cancer hospitals and institutions for crippled and disabled children (including congenital deformities and children crippled by poliomyelitis). Extensive follow-up programs were developed to determine the effectiveness of methods of treatment and the longevity of patients.

Care of tuberculosis and mental patients became of public responsibility primarily in state institutions located in isolated areas. In addition, some cities and counties developed similar institutions which frequently were subsidized by the state and had to meet state standards. The only other personal health care services were of a preventive or control nature in the area of communicable disease and many of these were paid for and operated by local units of government.

It was not until after the passage of the Social Security Act that states were able to initiate more extensive medical care programs, especially in the areas of crippled children, and the health of mothers and infants, from grants made available by the Federal Children's Bureau (Title VI Social Security Act) and required state matching funds. Venereal disease treatment did not become available until Surgeon General T. Parran of the U.S.P.H.S. launched his famous V.D. program during World War II.

Also during World War II, the Emergency Maternal and Infant Care Program precipitated the states into their first extensive medical care program for a select civilian population. States set standards for care, under guidelines set by the Children's Bureau, and developed payment formula for vendors - both

physicians and facilities. Based on prior crippled children's program experience, the EMIC program furthered the evolution of hospital and clinic standards, methods of cost-accounting in hospitals; and the administration methodology and financing of open-ended medical care programs. Although this program was temporary, it did lead to licensing of maternity and newborn infant nurseries and in some states to the licensing of hospitals.

Between the years 1920-60, other medical care programs were initiated. Cancer and cardiovascular (heart) clinics were established. Several states initiated hospital and ambulatory care programs for the rehabilitation of alcoholics. Special programs for transportation and hospital care of the premature infants were initiated. Most recently (1970-) programs in drug addiction and substance abuse have appeared. In the meantime, usually under auspices of an independent state department of mental health, community mental health and mental retardation programs began to expand, especially as federal and state funding became available.

Special population groups such as migrants and the socio-economically deprived persons have been receiving additional limited or even comprehensive health services in the past decade. Such services have been funded in a variety of ways, from several federal agencies, state and local governments as well as contributions from voluntary and private sources. They have been administered usually by some local agency, frequently the health department and coordinated by the state health department and the regional office of H.E.W.

There are rapid changes taking place in the financing organizatien of such ambulatory-primary care programs. Changes are also taking place in their organizational patterns and in the scope and nature of services offered. The future of these programs is dependent on continuation of federal subsidies and on the passage of national health insurance legislation. They do establish a second system of medical care for the indigent that is apart from care for middle income or "average" citizens. Hopefully, these units may become delivery centers in a single system of health care, or they may be eliminated as HMO's, regional health care systems or other devices for delivery of health care evolve. Their existence as they now are beyond the implementation of the national health insurance program is unlikely.

The licensing of hospitals, long-term care facilities and clinics began in the 1920's and 1930's. Initially, many of these programs were the responsibility of welfare and social service departments. By the mid 1930's these began to be transferred to state health departments. However, it was not until well after the passage of Medicare and Medicaid legislation and their requirement for the certification of facilities and subsequent legislation (1972) requiring certificates of need that health facilities were licensed by all the states. Many states have now passed legislation which sets standards for certificates of need for construction of new health facilities and for major reconstruction. Consequently, the rather ineffectual state hospital planning under the Hill-Burton Program has now become more realistic.

There has been a period of major organizational change in state governments in the past decade and it is continuing. These changes have affected the scope and extent of health programs and if the trend continues will bring about some theoretical consolidation of programs concerned with "human resources" but may result in disruption of health departments into new types of rearrangements of function and responsibilities. With the passage of the Social Security Act,

grants-in-aid became available to state health departments under Title V and VI. Later departments of mental health and some other units of state government also received health funds. The states formed an "Association of State and Territorial Health Authorities" in order to present a united front to federal bureaucrats. This organization reviewed all new administrative rules and regulations and proposed legislation as well as professional aspects of programs that were funded in part or to be funded by the USPHS or the Children's Bureau. In the beginning, that is mid 1930's and for many years thereafter, the ASTHA as well as the APHA passed resolutions and policies stating that the state health department should be independent as well as the health officer, a physician, who had direct access to the governor. Most state health departments were responsible for environmental health services, for crippled children services, the Hill-Burton Program and, a few, for community mental health programs. However, at no time were all health services of state government the responsibility of a single agency, and the hue and cry of public health professionals was for consolidation and against further splintering as new programs evolved. This splintering was occurring in the federal government, continued in most states and extended into local county and city governments. There were some notable exceptions; for example, the state of Maryland where Medicaid was placed in the health department. In most states medical care of welfare recipients was the responsibility of welfare departments, and when Medicaid was enacted,it was placed there.

The AMA and state medical societies fought against "socialized medicine" and did not want health departments to become involved in the delivery of general medical care for any group. Some public health leaders, including noted academicians (Haven Emerson),as well as national and state health administrators opposed placing medical care in departments of public health. There was no leadership in the USPHS toward the developing of a national policy of medical care delivery to the poor, nor for the development of a national health insurance program. The USPHS was not interested (other than its internal programs such as the Marine Hospital Service,Coast Guard and later the Indian Health Service) in developing any system of medical care delivery so as to make such services more available or accessible to people throughout the country, regardless of economic status.

Shortly after the creation of the Department of H.E.W. during the Eisenhower administration, there began to appear among the states a trend that more or less copied the organization of the federal government. But it was not until the 1970's that this trend to creat, in states, departments of human resources reached its climax. Most recent examples are Illinois, Massachusetts and Michigan.

A second equally important organizational and functional change which has been copied by many states is the creation of a separate Environmental Quality Protection Agency. This change resulted in removing many environmental health services from federal and state health agencies and placing them in a new setting, frequently lacking in any health or medical input. Thus, water, air and noise pollution, sewage disposal and solid waste handling are now removed from health agencies and made the responsibility of the new EQPA (EPA) where they are combined with natural resource, conservation and land use planning.

Thus by these two trends, creation of Human Resource Agencies and of separate EPA agencies, the role and responsibilities of state health departments (or divisions) changed. The health officer no longer has direct access to the governor and legislature, but reports to and works under an appointee of the governor, usually an individual who is responsive to the governor and his

political party. Combined in these departments are: welfare, public health, mental health, vocational rehabilitation, alcoholism, drug addiction, crippled children, all state institutions (except law enforcement and higher education), medicaid, special commissions or programs on aging, the blind, etc., boards of registration, and education. There are taking place reassignment of programs and responsibilities. Sometimes such changes take place whenever older department heads retire, resign or are replaced. This realignment means the disappearance of the traditional state departments of public health, mental health and welfare. Like functions and programs are being combined into units (some with prior names go with new titles). This licensing of health manpower (physicians, osteopaths, dentists, chiropractors, nurses, etc.) and health facilities (hospitals, long-term care facilities and clinics, etc.) together with certification for payment under Medicare and Medicaid and functions such as utilization review are placed in one unit (maybe the old health department) to which may be added the supervision of new types of health delivery facilities as these are developed under a national health insurance program. A second unit may be responsible for the management of all health care facilities operated by the state. A third unit may be charged with all purchasing of medical and health care institution services, home care, etc. from vendors on behalf of state beneficiaries. Liaison with local government and state aid to local health departments may remain with the old state health department (unit one) or be assigned with other inter-governmental functions in the health area to a fourth unit. This trend of realignment of services among units or departments within the overall state agency for human resources (or some similar title) is likely to accelerate as states are asked to assist in the evolution of health service delivery under some type of national health insurance.

D. The Role of Local Government and Community Health Services

Cities preceded any other government unit in developing health programs. Their efforts were two in nature: first, the control of pestilential (epidemic) diseases such as cholera, small pox, yellow fever, typhus, typhoid and malaria; second, the provision of direct medical services to the poor - usually limited and rendered by "city physicians". Cities' first programs were episodic, primarily in port cities and were characterized by the appointment of a temporary board to deal with the threat or actual occurrence of an epidemic. Gradually cities such as Boston, New York, Philadelphia, New Orleans and other sea ports, established permanent boards of health. By the last quarter of the 19th century these boards employed staffs of physicians and "sanitary inspectors". Their activities were expanded to new efforts concerned with the control of epidemic diseases - housing, control of nuisances - especially filth as a cause of disease, and instruction in healthful living habits. Concurrently, medical care of the poor began to require more extensive resources. Additional physicians, part-time and full-time, were employed. Pest-houses (Isolation Hospitals) were set up in then remote areas. In some cities, general hospitals were developed and operated by city governments. At first these city institutions were places where people went to die, or they where places for the sick-poor who could not be cared for at home. In other cities, voluntary organizations (non-profit) organized hospitals. Among the oldest and most notable of these are the Philadelphia General and the Massachusetts General in Boston. It was not until the onset of the 20th century that clinics for ambulatory care became an important part of public hospitals.

With the evolution of knowledge of bacteria, the practice of medicine and public health underwent phenomenal change. Cities evolved extensive programs of isolation and quarantine, immunization, and tuberculosis sanitaria (frequently subsidized by the state, etc.). Environmental sanitation programs, while still

concerned with nuisances and housing, now became important cogs in programs designed to control infectious diseases, especially in food sanitation (milk, meat, baked foods, food processing plants, restaurants), disposal of human and animal wastes, solid waste disposal, and the provision of a safe potable water supply.

In hospitals, emergency units were becoming more important and as the automobile became more prolific so did the demand on hospital emergency services. These demands were further increased during the middle of the 20th century with the continuing vanishing of the family doctor or the general practitioner. Current indications are that this trend in the increasing numbers of "primary physicians", physician assistants and the establishment of a variety of groups of pracitioners who will furnish not only primary care but continuing comprehensive as well as preventive health services.

The hospital, toward the end of the first quarter of the 20th century, began a metamorphosis. It was no longer a place where the sick-poor went to die; the hospital was becoming a place of hope - where the sick (poor, middle class and the rich) went to recover from an illness that could be treated better in a hospital setting. By the middle of the 20th century, hospitals became places to maintain health as well as to receive care for illness or injury. The rapid evolution of pre-paid hospital insurance has led to over-hospitalization or over-utilization of hospitals. Currently, there is a trend to reduce hospital over-utilization by evolving systems of financing and systems of health care delivery that encourage health maintenance, preventive services and ambulatory care. These trends are brought about by the escalating hospital costs and in an effort to develop reasonable cost-containment of medical care costs.

Another series of trends can be observed in hospitals as they relate to public health. Among these are their efforts to become involved in extra-mural programs - educational, public relations, operation of satellite ambulatory and/or in patient facilities, the involvement of consumers in governing bodies, the training of allied health professions, the evolution of home care programs, and their role in the blue cross programs (to mention only the more important).

Counties, on the other hand, were the last among governmental levels to develop health programs. In 1911, the first rural county health departments were organized, one in Yakima County, Washington, as a result of a typhoid fever epidemic, and the second in Guilford County, North Carolina because of concern for the health of school children. The Rockerfeller Foundation's concern for hook worm in the South, the W. K. Kellogg's concern for community health in southwestern Michigan, the interest of the Children's Fund of Michigan in the dental health of children in central Michigan and the interest of the Milbank Memorial Fund in tuberculosis control in Cattaraugus County, New York helped to initiate the development of county health departments in the 1915-1935 period. Thereafter, grants under Titles V and VI of the Social Security Act and state-aid assisted in the development of county health departments and later multiple county health departments.

Haven Emerson, working under the aegis of the American Public Health Association and after consultation with state health officers and their associates, developed a plan entitled "Local Health Units for the Nation". Published as a Report of the Sub-Committee on Administrative Practice of the APHA in 1945 (Commonwealth Fund), this document set a pattern which defined the basic five

service areas, staffing parents and geographic basis for local health units. A sixth basic service (chronic disease control) was added a few years later. The purpose of this report was to set a minimum standard and to encourage total coverage of the nation. For each state there was a suggested pattern of covering its entire area with single or multiple county units, each employing as a minimum the basic team of a health officer, public health nurse, a sanitarian, and a secretary. Personnel population ratios were included; these were developed as minimums and were derived on concensus from the informed expert opinions of members of the subcommittee. This report together with further reports of the APHA on "Services and Responsibilities of Local Health Departments" (three) have shown the gradual evolution of the basic philosophy as to the role of local health departments. From a traditional small unit concerned primarily with the control of communicable diseases, the promotion of maternal and child health, limited environmental health services, health education, and supporting statistical and laboratory services, local health departments are now envisioned as offering leadership to all community health services, including health data systems, identification of health needs, resources and manpower; planning, execution and evaluation of comprehensive programs of health; health education and maintenance; preventive, therapeutic, and rehabilitative services that are continuing and comprehensive for special groups of the population (including the socio-economic disadvantaged, migrants and areas without health resources, etc.); ambulatory care; home care; and, in some instances, acute general hospital services; special and chronic hospital services; long-term facility care or the supervision of such facilities; and the coordination of its programs and that of voluntary health agencies, hospitals, other government agencies and private enterprise.

As no two states have similar health services delivery patterns, no two cities and counties have identical programs. Generally, within a state there are, because of state laws, regulations and financing, similar patterns of organization and operation. Even within a state there are great differences in expenditure of dollars per capita, scope and depth of services offered, population-staff ratios, training of personnel, consumer involvement, etc.

Now, thirty years after Haven Emerson's work was carried out, there are some states without any local health departments (Vermont), some which have a few large city health departments (Maine, Alaska), others which have a few rural and a few city units (Minnesota), some that have most of the state covered (New York, Illinois), some that have nominal coverage of the entire state (particularly in the South), and a very few states which have complete coverage with local health programs that are conducting a reasonable program of basic services (California, Florida, Michigan). A few states deliver local community health services through a pattern of state, regional or district offices (Rhode Island), others combine local health units with state regional offices (Ohio, Massachusetts, New York).

The independence or autonomy of a local health unit from the state health department varies completely from state to state. In every state, local health units are required to enforce state laws and regulations. In some states, the local unit is, in fact, a branch office of the state (Maryland); in others, it is a subsidiary unit of the state with some flexibility as to add-on programs (Kentucky, Florida). As the state makes financial assistance available, it sets standards for performance and personnel. Such standards may be loosely enforced (Illinois) or strictly enforced (Florida). Federal funds may (Michigan) or may

not (New York) reach local health units through redistribution of federal grants-in-aid by the state. All local health units may seek direct federal project grants, but these are processed through the state health department and comprehensive health planning A and B agencies. In a few states, the local health department is quite autonomous except for meeting state standards (Michigan, New York).

Innovative programs of organizing and financing local health services have been developing. California has the ability to contract with smaller counties to deliver public health services. Counties may contract with each other for specific services, to employ a health officer, and to employ other health personnel. They may merge or become associated boards and in some states form permanent districts or regions. In one state (Michigan), a state university contracts with a county board of health, or a district, to provide supervision and to share costs in exchange for teaching facilities. Health, mental health and/or welfare may be combined in cities or counties. Public hospitals may be separate or a responsibility of the health department. Home care services, long-term facilities, ambulatory clinics, and neighborhood centers are other examples of some health department programs. In Utah, for example, the state has been divided into regions. In each region, there is to be a multiple county unit combining health, mental health, welfare and all other human service agencies in a single governmental unit that replacates the state organization. It is financed jointly by the several counties and the state.

The national trend to merge or combine local health units which serve small populations has been rapidly accelerating. Research as well as experience have indicated that a health department serving a population of 200,000 to 250,000 has a more efficient operation, can be more innovative, can meet emergencies better, has a more progressive program, can attract and hold better qualified personnel, is able to deliver a wider scope of personal and environmental health services, is able to seek and obtain outside financial support, has a better reputation for quality of services, is more likely to involve consumers, has more adequate public support, and plays a more important role in the planning, delivery, evaluation and coordination of comprehensive health services.

In summary, the role of local government (city, county and regional) varies greatly from state to state and within a state. In some states, local government has a nominal or negligible role in health. In others, the role of local government has been expanding from traditional environmental health, health maintenance, and preventive personal health services to the administration of health facilities and manpower in delivery of therapeutic services, including:

1. tuberculosis ambulatory services
2. administration or purchase of tuberculosis institutional care
3. venereal disease ambulatory care (or purchase from vendors)
4. purchase (and frequently operation) of institutional care for communicable diseases
5. well infant and child clinics
6. prenatal clinics and expectant parent programs
7. school children (and personnel) health services
8. home care services - including home maker services, nursing care, physical and occupational therapy, social work, and, less frequently, nutrition and dental services
9. family planning ambulatory services
10. dental clinics

11. alcoholism ambulatory care (rarely institutional services)
12. drug rehabilitation services
13. screening programs; for adults and Medicaid Eligible Children
14. special clinics; for oldsters, hypertension, diabetes, etc.
15. comprehensive ambulatory health services for socio-economic disadvantaged, special groups such as migrants, civil service employees, and prisoners
16. administration of general hospitals - usually only in large metropolitan cities or counties and these may no longer be a responsibility of the health department
17. administration of long-term care facilities (as in 16)
18. administration of special health facilities, for communicable diseases, tuberculosis, mental patients
19. administration of a community mental health program
20. provision of clinical laboratory supportive services
21. special programs such as lead poisoning, sickle-cell anemia, etc.
22. licensing, certification review, and/or supervision of health facilities, inpatient, outpatient, and home care
23. classification of patients in long-term care facilities for payment under Medicaid
24. Participation in utilization review

There are some local health departments that provide staid conservative programs - primarily holding operations that may be mobilized to go beyond the past only at times of a health emergency. Other departments are moving forward; the only limitations being in the availability of resources and manpower.

Some of the most recent areas of innovation are:

25. participation in research of comprehensive health care delivery systems
26. provision of emergency medical and health services including provision or supervision of ambulances
27. cancer, cardiovascular, smoking, kidney disease (including dialysis and transplant)
28. coordination of total community health programs, including the promotion of new programs and extension of existing programs
29. programs designed to give public accountability and understanding
30. cost-containment and programs to evaluate effectiveness and efficiency of health programs
31. An increasing role in health policy decision-making and planning in partnership with labor, commerce, industry, private health enterprise, other health agencies and the consumer
32. Motivation of persons and agencies to develop proper attitudes and health values and to seek appropriate health care measures
33. training of allied health and health personnel

More and more local health departments are becoming involved in an expanded role in the health services delivery system. Unfortunately, many areas of the country are without local health units; some local health units are ultraconservative and lack either or both agressive leadership and essential resources. The financial support, in almost all states, of local health departments is the real estate tax (either a special health levy or a portion of the general tax). This

is supplemented by "general revenue" available to the city or county (most recently a portion of federal tax sharing) and perhaps some state-aid and federal grant-in-aid funds (both of which are usually small), fees, and miscellaneous gifts. The tax support base is too small and limited to support an expanding health program (especially if expensive direct personal health services - medical, dental, and institutional - are included). Health departments, state and local, will require extensive financing from a national health insurance program if they are to carry out functions and activities that may be assigned to them.

E. The Future Role of Government in Health

No one can accurately predict the future. Many of the respondents to our letter of inquiry were adamant on this point. Of over 60 respondents, almost all were of the opinion that the future of the health care polysystems in this country will be determined by federal legislation. There was universal agreement that:

1. There is a distinct probability that the role of the federal government in health will increase at an accelerated pace within the next decade.
2. Federal expenditures for both personal and environmental health will continue to increase annually, in part due to rising costs, in part to an increased number of beneficiaries and in part to an extension of benefits.
3. Federal legislation will be enacted within the next decade to increase the number of beneficiaries and to increase benefits under federally sponsored health programs.
4. Some form of "national health insurance" will be enacted in the next few years, and after a tooling-in period, implementation will start in 2-3 years thereafter.
5. It is probable that some form of health maintenance and preventive services will be provided in an effort to keep curative (therapeutic) medical costs down.
6. Cost-containment provisions may require some changes in the health delivery systems and practices. Some of these may cause adverse criticisms of both practitioners and institutions that are vendors.
7. An effort will be made to extend the government's role in supervision in an attempt to exert some control on quality and quantity of care - both by practitioners and facilities.

Other trends which were pointed out by some of the respondents, persons interviewed, and readings are:

1. Decrease of federal government's support of basic research in the health area and diversion of these funds to applied research in the delivery of health services.
2. A continuing lessening in the federal government's support of training in research and top health professions.
3. Funding of HMO and other experimental models for delivery of comprehensive continuing health care systems.

Although there were differences as to the future role of local health departments, only a few saw a lessening role, with personal health services removed to a new agency and environmental health services diminished or transferred to a new local EPA. The vast majority foresee an expanding role for local government in the health care delivery system.

Many visualize a regional organization, multi-county. This organization may or may not be a health department. Some form of health maintenance, preventive, therapeutic, rehabilitative - continuing and comprehensive - services will have to be universally available on an ambulatory, inpatient, home care and long-term care unified system approach. Free standing, hospital associated HMO, Medical Foundations, group practices and the like are being gradually developed in an attempt to deliver these services on a capitation or fee for service financial basis. Many are developing as independent pre-paid insurance units where premiums are to be actually based on experience. Some are non-profit; others are profit. Some are privately organized; others are public. It is not forecast that such units (HMO, Medical Foundations, etc.) will be able to provide total coverage for all of the population. Central city or core areas of metropolitan cities, sparsely settled rural areas, communities without adequate doctors and/or hospitals, etc., special population groups such as migrants, the minorities, the socio-economically disadvantaged, and the aged will not be covered by voluntarily organized units. Populations in some of these areas are high risks and cannot be included in such units at standard capitation or uniform cost formulae. Moreover, there may not be enough health professionals or facilities to enable the independent setting-up of new health delivery systems - the financial risk is too high and would not attract physicians.

Hence, if we are not to repeat the errors and experience of Title XVIII and XIX wherein home health services and even long-term care benefits were made available but the system did not exist to see that these services became available, then government must evolve a system whereby eligible beneficiaries will have reasonable access to services provided by the program. This need may necessitate the furnishing of transportation of patients to not-too-distant health resources. But in addition it will require the development of at least primary care for entry into the system and in some areas even more comprehensive health resource facilities.

Some of our associates visualized that the future role of local health departments may include the responsibilities as follows:

1. Promote the development of primary care centers
2. Develop and operate primary care centers until they are taken over by independent groups
3. Continued operation of primary care centers for general populations, special groups as migrants, socio-economic disadvantaged, etc.
4. Supervise the operation of primary health centers in their region and coordinate the health care systems, including utilization review, inspection for certification, identification of unmet needs, planning and policy decision roles in evolving added services and programs, evaluations, health information system management and evaluation.

This changing role of the local health department has taken place in Montreal, following the Quebec Provincial Health Insurance Program enactment. The Montreal Health Department organized personal health delivery units for population groups not served or included in privately organized units. The department plans to continue to operate these units until they are taken over hopefully by a private group.

As several of our consultants pointed out, the ideal unit of government to participate in the future health care delivery system is regional. One in which all "human service" elements of government are combined. Thereby lowering administrative costs, avoiding duplication, coordinating for the total care of the

individual or family and finally making entry into the system convenient, humane, acceptable, readily available and accessible. Larger regions would have decentralized branches. Such a regional unit would be responsive to the consumers, to local officials, could coordinate state services and receive state support and be the regional authority for federal health and welfare programs. Since the unit would look after the total needs of the individual, it should have responsibilities in the field of all human resource activities including public health, mental health, environmental health, emergency medical services, health maintenance, preventive, therapeutic, rehabilitative services, ambulatory home care and institutional as well as social services and counseling.

Two plans have been developed along this line: 1) Illinois - in the "Illinois Community Health Services Study" 1971. The creation of regional Human Resource Areas was proposed as the ultimate goal of the reorganization and improvement of local health services by 1975-1985. The plan called for the assignment of state service personnel from the central and regional offices of state agencies to the local AHRA (Area Human Resource Agency) which was under the control of representatives of each of the counties in the area including consumers, practitioners and public officials. The consumer would be enabled to go to the AHRA office or, in large areas, to a branch office, regardless of what his need might be in the total health-social welfare field. Here, there would be a combination of state and local services assuring prompt attention to the individual's needs with referral within the organization rather than to another unit of government which usually would be at some other location.

Early entry into the system would be facilitated by a trained and experienced "intake worker" and by the periodic issuance of an identification card. The plan calls for the integration of all health services (state and local) including tuberculosis, public health, mental health, crippled children, vocational rehabilitation, emergency medical services, substance abuse, alcoholism, environmental health services, together with family services, child care services, other social welfare services, commission on aging, comprehensive health planning and counseling. This unit could contract, deliver, and or supervise the delivery of health maintenance, screening and diagnostic programs, preventive, therapeutic, rehabilitative programs on an ambulatory, home care, hospital, or long-term care facility basis. In addition, it would integrate health and social programs. It would have a role in the training of health, allied health and social welfare personnel. It could participate in applied research and innovative health services delivery systems. Since this plan has been proposed, there have been several changes in the Illinois Public Health Program, but as yet the plan has not been implemented. 2) The state of Utah has developed a regional human resource regionalization program. Implementation has started, and the entire state is to be covered in the next two or three years. A region is comprised of one or more counties, each of which helps to support the staff and the facilities. The state contributes funds to employ the top leadership and management staff. The region is to gradually expand its fields of activity from public health to the entire gamut of human resources as organized in the state.

No doubt there are other examples of the integration or coordination of health, mental health, welfare and institutional care such as San Mateo County, California, which has been a leader in this type of coordination for over a score of years. Several other local departments in California are combining - Los Angeles (city and county) Health, Mental Health and Institutions.

Fortunately, the organization of local government and state government is neither static nor inflexible. There is ample evidence that state and local government organizations can be adapted and modified to meet new responsibilities and new programs. Thus, when national health insurance is passed, if state and local governments are assigned specific responsibilities, they are capable of tooling-up to accept the new tasks if adequate time and resources including manpower are made available. No doubt there will be a variety of ways in which states, cities and counties will respond. Hopefully, they will be given sufficient latitude to innovate, experiment and demonstrate a variety of approaches to meet the needs and demands of the new programs or any others that may evolve.

The final question, therefore, is not what is the future role of state and local health departments. The question is: "What is the future role of state and local government in the developing health programs of the next two decades." Our conclusion is that the federal government must either evolve an entire hierarchy of regional units to deliver, manage and supervise the delivery of health services, or it may utilize, more or less, existing state and local (regional) units to carry out some or all of these responsibilities. The most probable course is a middle one in which some functions and responsibilities will be retained by federal government decentralized units while others may be delegated, contracted, or otherwise assigned to qualified state, local or regional units of government. Not all areas of the country would initially have or be able to develop adequate regional units to coordinate local, state and federal health and social welfare activities. Thus, there can be expected provision in the legislation which would permit the federal agency which is responsible for the health insurance program and/or the delivery of health services to adapt its program in a way which will most effectively and efficiently meet the requirement and utilize the existing resources of a given region.

There is no doubt in our minds that some state and local health departments will shun such a program as though it were a plague. Others will promote their interest in participation. Some will jump-on-the-bandwagon and go along. Eventually, most, if not all, health departments will pitch into the program when they perceive that the alternative is the creation of another local (region) agency and their own gradual dissolution. (A case to illustrate this - In 1933, the Massachusetts Department of Public Health, with the support of the Massachusetts Medical Society, had decided not to participate in the E.M.I.C. program. In May of that year, a new state health officer was appointed and announced that Massachusetts could not afford to be the only state in the country that did not provide the benefits of the program to the wives and children of servicemen.)

In conclusion, the future role of state, local and regional government in the national health program is an expanding one. While the organizational pattern can be expected to be considerably altered, present state, local and regional health departments will be absorbed in the new health delivery system - or cease to exist.

IV. Training Health Manpower For This Role

It was the concensus that there is a shortage of trained public health personnel now and that the shortage will be much greater as health programs expand in the next ten years. The shortage was the most acute in top management people - especially physicians - who could become directors of programs and agencies. Several respondents indicated the increasing need for schools of public health to train non-physicians to assume positions as program directors and local health officers. Others pointed out the need for leadership in public health nursing, health education and sanitation. It would appear that the need for trained public health personnel may be even greater in voluntary agencies - especially for agency directors and directors of health education programs. There is little question that in the next ten years public health manpower needs will exceed the supply, especially if there are cutbacks as may result from the federal economy wave.

Students should be prepared to meet the needs of positions they are to assume on graduation and to be able to meet the needs of positions that they may hold within the decade after graduation. Since it is impossible to predict what the needs of personnel may be ten years from now, it is important that the student's training provide him with attitudes, skills, and knowledges he can utilize in adapting his work production to future demands.

Perhaps the most important facet that he should acquire is to learn "how to learn" (as one of the respondents put it). He must be constantly learning on the job as well as in professional meetings, reading, associations with peers and attending short continuation or extension courses. Some personnel should return for further formal education in which theory and practice are "combined and intertwined". Every graduate should have instilled into him the concept that graduation is only the beginning and that learning is a continuous process throughout his professional career.

Most respondents felt that schools of public health were an appropriate place where public health personnel should be trained. Many indicated that there should be greater opportunity for the student to obtain some teaching from other faculties such as schools of business administration, public administration, and professional faculties as in medicine, social work and denistry or nursing.

There was evidenced a loud call for less theory and more "practicum" - not only for students but for faculty. Several recommended that in addition to developing teaching facilities in cooperation with health service agencies, that staffs of agencies serve on faculties of schools of public health and that faculties of schools periodically be assigned to operating agencies so as to have a more realistic understanding of the problems and needs of health agency personnel. A few indicated that the practice should comprise as much as 40% or 50% of the total school time. Others pointed out that there is an obvious change in the composition of the student bodies of schools of public health. In the past, most students were mid-career and had some or extensive work experience with a health agency. This experience was a source of information to other students in the exchanges that took place in classroom and other discussions. These students may have learned as much or even more from each other than from the faculty. Now the majority of students enter directly on graduation from college with no or little relevant experience. Hence, students are less likely to learn health-related material from each other. Moreover, it was pointed out by several respondents that faculties, too, have changed markedly. Formerly, many if not most

of the faculty were physicians with extensive public health backgrounds of work experience, teaching and research in the health fields. The present faculties of schools of public health are multidisciplinary. Many are recent Ph.D.'s with no experience in the health industry and some with little knowledge of the health sciences. Thus, students are not able to learn health relevant material from such faculty. Assignment of new health career students to field placement with actual work experience becomes more important if they attend schools of public health where faculties lack knowledge and/or work experience in the health industry.

For mid-career students, a different track was suggested; namely, less practicum in schools and more theory and an opportunity to relate the two in seminars that are oriented to actual problems experienced in their prior work situation.

A few respondents thought that schools of public health were spreading themselves too thin by attempting to train persons in too many program areas. The concensus seemed to be that generalists should be the goal of a one-year M.P.H. program. Each graduate should have adequate instruction in the health sciences and since he would take a position of management that he should receive instruction in administration and the other essentials of management, such as fiscal and personnel management, planning operation, evaluation and reporting. Several included such areas as concentration on the "service aspects" and emphasis on "public accountability".

It was further pointed out that graduates of a specific program in a school of public health frequently accept a position in another area of health. This was especially true of graduates of the several programs in administration; namely, public health administration, medical care administration, hospital administration and health planning. Furthermore, they frequently shift in the next few years to other types of health services administration. A suggested solution that was offered is as follows: Offer all health administration students competency in health science and a knowledge of health organizations and operations so as to enable them to manage any type of health agency. If some students wish to concentrate in a special area, this could be done by elective courses or subsequent study.

Several letters pointed out the need to have other types of specialists, such as material and child health, public health nursing, environmental health, etc., have adequate instruction in public health administration. Many of these specialists become program or division administrators and lack management skills; some become agency chiefs and lack in addition the skills, knowledge and attitudes required in this top leadership position. As indicated elsewhere, several respondents were of the opinion that all graduates should be firmly grounded in health science and management.

Many replies called attention to the need of health agency involving consumers, private enterprise in the health industry, labor, commerce, industry and public officials. It was indicated that graduates did not know how to realistically bring the consumer and others into true partnership in the decision making and policy formation roles. Schools of public health should develop expertise and seek actual practitioners with skills and know-how in this area to teach students real and not just nominal consumer involvement methods, techniques and skills.

To the query as to the adequacy of a one-year M.P.H. program, the vast majority said, "Yes". A few suggested that some graduates were: 1) not prepared to do budgeting, or prepare a program or project proposal, 2) graduates did not know how to administer or supervise personnel, draw up job descriptions, or prepare merit reports on staff, and 3) they were not competent to deal with employee grivances or to deal with unions. A few suggested that a two-year program was desirable since it would prepare graduates more adequately; but, most indicated they could not give a two-year educational leave to an employee and that it would be an impossible financial burden on the student and his family as well as on the agency. Hence, they did not want a two-year M.P.H. program for mid-career students. A few health officers suggested that physicians should attend schools of public health only after public health experience. They could then contribute more to other students, see more relevance in what is taught and possibly could finish in two semesters.

One health officer, an alumnus, with over two years of experience prior to attending the University of Michigan School of Public Health, thought that he learned very little that was useful during his M.P.H. course. Several persons in a single state health department were very adamant in their opinion that the University of Michigan School of Public Health graduates in environmental health were not any better as a result of obtaining their M.P.H.

Most of the respondents in operating agencies wished that the schools of public health teaching was more relevant to current and future needs. They expressed their opinion that the faculty was uninformed of the many changes that were taking place constantly in the daily tasks of health departments. Several recommended a variety of ways of meeting this problem:

1. Develop a joint curriculum committee or advisory committee with representatives from service agencies
2. Include agency staffs on faculties
3. Assign faculty to service agencies, periodically for six months or a year
4. Employ more faculty who had a firm base of public health experience
5. Develop teaching and research programs in health agencies
6. Place students for continuous and extended periods with service agencies
7. For students who entered school without public health experience require a residency in the field.

Another area that received attention was the problem of education of all types for employees of health departments while they were on the job. Several needs were expressed:

1. Much more expanded short course offering (non-credit) on campus and off campus. Planning of such courses should be a joint responsibility of representatives of the service agencies and the faculty.
2. Structured longer courses - some for employees without a bachelor's degree, some for post-bachelor's - whereby employees could obtain credit towards a master's degree and the eventual development of the total M.P.H. offerings on an off-campus basis.
3. From two states, suggestions were made to develop off-campus and on-campus extended weekend course offerings so that over a period of several years (two or three) students could obtain their M.P.H. degree. Such offerings could take place on Friday and Saturday and be offered on a regional basis.

4. There is an outstanding need for prebachelor's education for persons with associate degrees in nursing and sanitation. On the job, off-campus teaching, leading to a bachelor's degree in public health nursing and in sanitation were suggested as an area for concern of schools of public health.
5. Post-master's course - some refresher in type, some in new areas, short courses, seminars, conferences and a whole mix of jointly planned, in-service education both on and off campus.

Several respondents pointed out that the greatest need now and in the future will be for persons who can fill leadership roles at top and middle levels of management. Schcols of public health should concentrate on students who in due time would be able to assume these leadership roles rather than on the lower level of supervisory and service personnel. Some suggestions as to what these leaders should be able to do are:

1. Coordinate medical services into a more unified system emphasizing health maintenance, preventive care and delivery of primary care
2. Evolve and supervise contractual relationships with the private sector of physician and health facilities
3. Coordinate health and welfare services
4. Continue management of such programs as:
   a. Communicable Disease Control
   b. Environmental protection - nuisances, housing, industrial and community
   c. Consumer protection
   d. Programs for the aged
   e. Programs for the chronically ill
   f. Provide and/or supervise health services delivery systems
   g. Mental and emotional health
   h. Child abuse and abuse of the elderly
   i. Abuse of drugs and alcohol
   j. Provide or see that provision is made for adequate care of the mentally defective and deficient
   k. Analysis of data and identification of health needs
   l. Planning and evaluation
   m. Provision of emergency health care
   n. Develop branch offices and outreach activities
   o. Motivate individuals and families to seek care
   p. Develop programs to support small town and rural practitioners
   q. Provide public health leadership and know-how for the development of new health programs.

Specifically some suggested that the curriculum include more on:

1. law
2. political science
3. sociology
4. statistical aspects and application of computers to health information systems - data analysis and interpretation for planning and evaluation
5. policy-decision making
6. economics and fiscal management
7. personnel management
8. community organization
9. administration and management
10. epidemiology
11. marketing research

12. a thorough knowledge and understanding of health science and the organization and operation of health systems
13. methods of motivating action in individuals, group, practitioners and communities.

One respondent put it this way - the graduate should be a "human service worker" who is concerned with the "quality of life". Another suggested that the schools of public health bring together a "think tank" to assist in developing the school curriculum (modeled after the governmental or industrial pattern). Several respondents strayed from the direct mission of the schools of public health and pointed out the needs for all types of health, para-medical, allied, health-related personnel. Some thought schools of public health should assist in setting up programs to train physician-assistants, allied health and other health-related personnel. One or two suggested that the time has come to reconsider the desirability of developing bachelor's programs for sanitarians, health administrators, public health nurses, health educators and others.

Finally, some simply stated that students should be taught all of the things (knowledge, skills and attitudes) mentioned in the letter of inquiry. One respondent was of the opinion that many just-graduated bachelor's degree people who were entering the school where attracted by the benefits of a traineeship, especially in a poor job market, that they did not have a real interest in public health, that a good portion of these never entered public health or did not stay in health jobs. A review, conducted periodically, of alumni as to their employment status over five or ten years would help to determine the accuracy of this criticism.

In summary, the staffs of government health agencies in the future will need top and middle leadership that is competent to deal with management problems in the total health care system. Their training should prepare them to take beginning positions of management in health agencies in which they could evolve to more responsible positions and at the same time continue their education throughout their career. They should be taught "how to learn", be prepared to be change agents, should be knowledgeable in community organization including the identification of community leaders, how to be democratic leaders, how to work as a team member with staff, other groups and coordinate programs that comprise the community health systems.

They should not only be leaders, but know how to make decisions and obtain favorable decisions from associates and superiors. They should know interpersonal relations, be knowledgeable of psycho-social factors and thoroughly educated in health sciences and the organization and operation of government, community and private sectors on the one hand, and health systems on the other.

Other specifications are: adequate base in epidemiology; vital statistics; health information system management including data collection, analysis, interpretation and retrival; application of computer science; fiscal management, including budgeting, allocation, encumbrance, accountability, cost-benefit analysis, efficiency studies and ability to develop project and program proposals; personnel management including recruitment, merit system operations, job descriptions, retention, supervision, promotion, discipline, separation, records, grievance procedures, union negotiation, staff morale, payroll and fringe benefit management, travel, liability, etc.; management - top, middle and low including supervision, assignment of responsibility and authority, reporting and review methodology,

obtaining and assigning resources; handling of operational problems and emergencies, equipment and supply management, inventory and control and maintenance.

Among the most important assets of a public health leader is a humbleness and a readiness to seek help and advise, to work as a member of a team, to work with rather than on or for people and to be ready to change if the need indicates.

FOOTNOTES

1. Commission on Education for Health Administration, Future Trends in Health Care Delivery: A Forecast, 1973.

2. Getting, Vlado A., Illinois Community Health Services Study, 1971.

3. Getting, Vlado A., Report of Sabbatical Studies and Activities, September 1, 1971, through February 29, 1972, 1972.

4. The Changing World of Work, Report of the Forty-third American Assembly; November 1-4, 1973; Arden House, Harriman, New York, The American Assembly, Columbia University.

APPENDIX A

Please return to: Vlado A. Getting, M.D., Dr.P.H.
School of Public Health
The University of Michigan
Ann Arbor, Michigan 48104

May 24, 1972

1. What are the requirements for a local health officer in your state?

	City		County	
	Yes	No	Yes	No
None	___	___	___	___
M.D.	___	___	___	___
D.O.	___	___	___	___
M.P.H.	___	___	___	___
Other (Please Specify) _____	___	___	___	___
Is Experience Required (Number of Years___)	___	___	___	___
Is State Approval Required	___	___	___	___

2. Does the requirement differ for large and small departments? Yes___ No___
(If no, go on to question 3)

What is the requirement for small departments?
Population size under _____

What is the requirement for medium-sized departments?
State population range _____

What is the requirement for large departments?
State population size over _____

3. Are the requirements different for cities, counties, multiple county units, and state districts or local regions?

City

County

Multiple County Unit

State District

Local Region

4. Who sets the requirements?

	STATE		LOCAL	
	Law	Regulation	Ordinance	Regulation
	_____	_____	_____	_____

5. Are there any non-physician health officers in your state? Yes____ No____
(If no, go on to question 6)

How many? _____ Do they meet qualifications? All_____

Some_____

None_____

If some or none, explain.

6. Have the requirements been changed in 1970? Yes ____ No____

In 1971? Yes____ No____ In 1972? Yes____ No____

Please attach a copy of all laws, regulations pertaining to the appointment of a local health officer.

7. Please attach copy of anticipated changes.

8. Are non-physician graduates of schools of public health with an M.P.H. or equivalent being employed in state agencies in positions as:
(If no, go on to question 12.)

	Yes	No
Program Directors	____	____
Administrators	____	____
Deputies	____	____
Planners	____	____
Other Responsible Decision-Making Positions Specify_____	____	____

9. How many and in what positions where they employed? Please enter number.

	1969	1970	1971	1972 (1st 4 mos.
Program Director	____	____	____	____
Administrator (Describe)	____	____	____	____
Deputy (Describe)	____	____	____	____
Planner	____	____	____	____
Other (Describe)	____	____	____	____

Please attach job descriptions for these positions.

10. Are you finding adequate numbers of qualified applicants (non-medical M.P.H. graduates or equivalent)? Yes____ No____

	Now	1973	1974	1975
No Vacancies Anticipated	____	____	____	____
Local Health Officers	____	____	____	____
Program Directors	____	____	____	____
Deputies	____	____	____	____
Planners	____	____	____	____
Others	____	____	____	____

11. At the University of Michigan, the present program for candidates for a M.P.H. degree in Public Health Administration calls for three semesters of study with a minimum of 36 credit hours of course work (six of which may be in field experience) for persons with a bachelor's degree and adequate public health experience (usually three years). Candidates with no experience are required to take five semesters of study with 60 credit hours.

Are these graduates adequately meeting the responsibility of their positions?
Yes____ No____ If no, explain how they fail and what more is needed.

12. A new plan of study for Public Health Administrators (not physicians) would require candidates to have a bachelor's degree, adequate public health experience (1-3 years depending upon type and degree of administrative responsibility), four semesters of academic course work with a minimum of 48 credit hours, and a supervised residency program of 8 months equivalent to an additional 24 credit hours. These graduates would be prepared to assume responsible public health administration positions upon graduation.

Would this type of graduate be more acceptable in your state?
Yes ____ No____ If yes, in what positions?

13. When graduates are looking for positions in your state, where should they apply?

Name_____

Agency_____

Street_____

City_____ State_____ Zip_____

Telephone: Area Code_____ Number_____

Name and title of person completing questionnaire:

_____

_____

Address_____

_____

Telephone: Area Code_____ Number_____

Date_____

APPENDIX B

September 20, 1973

Dear Colleague:

I have been asked by our School to prepare a position paper on the likely functions of official health agencies in 1980 and the role of schools of public health in preparing people for these functions.

With the advent of comprehensive health planning, some form of national health insurance, great advances in the technology of medicine and health, it is likely that the role of state and local health agencies will change vitally. The position paper on which I am working will be an attempt to forecast this by updating probable lines of development. It is to help clarify this that I ask your help.

Have you or your associates made an analysis of social, economic, technical, or other developments which are likely to change the work, responsibilities, functions and organization of either governmental or voluntary health organizations? Have you prepared a 5 or 10 year plan for your agency? Have you projected the changes which are likely to take place in health agencies which deliver both personal and environmental health services? Have you developed a model program or organization which you believe will be most able to deliver the health programs of the future?

This series of questions -- as to what is the future of official and voluntary health agencies -- leads to a second series of questions relative to the manpower needs to bring about the necessary changes and then implement the "new model".

What are the health agency manpower needs of the next one or two decades? What kinds of people (types of personnel) will be needed, how many, what training and experience should they have?

How should health agency personnel of the future be trained? What should they know of technology, health sciences, social sciences, administration, planning, management, supervision, evaluation, computer science, etc.? What facts, attitudes, value judgments, skills should be included in their training?

Who should train health agency professionals of the future? How long should the training be? Should it be primarily theoretical or experiential or a combination and if so in what proportion?

Are schools of public health the places to offer all, some, or none of this training? Are other types of institutions of higher learning better qualified to train health agency professionals? What are your reasons for holding your view?

What should be the tie-in, if any, between educational institutions and operating agencies -- for students -- for faculty?

It is our intent to prepare a report that will encompass these and related questions. The report must be completed during November. Hopefully, you will assist us by contributing some of your ideas, thoughts, as well as sending us any plans, documents, papers or notes which you, your associates, or your institution have developed along these lines.

Rather than developing a lengthy questionnaire we are sending you this informal request for your much needed assistance. Please let us have your reply at as early a date as possible. If there are others in your area who may be of assistance to us, please share this letter with them or send us their names and addresses. Thank you in advance for your assistance.

Sincerely,

Vlado A. Getting

Vlado A. Getting, M.D., Dr.P.H.
Professor of Public Health
Administration

VAG/jap

UNIVERSITY OF SOUTH FLORIDA
TAMPA • ST. PETERSBURG

COLLEGE OF SOCIAL AND BEHAVIORAL SCIENCES
DEPARTMENT OF ANTHROPOLOGY
TAMPA, FLORIDA 33620

813: 974-2138
813: 974-2140

December 16, 1974

William Hill, M.D.
Polk County Health Officer
Winter Haven, Florida 33880

Dear Dr. Hill:

Doctors Northcutt and Shiloh, and I wish to express our appreciation for your attendance at the Advisory Committee meeting for the carcinogenic survey. We believe that your suggestions on various methodologies have improved the potential value of this study and are quite eagar to begin.

I have delayed sending a copy of the minutes in the hope that the National Cancer Institute would have sent its list of carcinogenic agents and I could include the list, questionnaire, and cover letter for your inspection. Unfortunately, we still have not received the list and are presently engaged in discussions about the questionnaire with the Florida Regional Medical Program, Inc. I will send this information to you as soon as we have the material. Copies of the quarterly report are being made at Florida Regional Medical Program, Inc. and will be forwarded to you under separate cover at a later date.

Thank you again for your time and effort put forth on this project.

Sincerely,

Robert I. Gilbert, Jr.
Research Associate

RIG/gb

Encl.

SCHEDULE OF PROGRAM ACTIVITIES

1.0 DURING THE PERIOD OF OCTOBER 15, 1975 TO OCTOBER 31, 1975

1.1 Delineate those industries in the four county area (Hillsborough, Pinellas, Polk, and Manatee) to be selected for personal interviews regarding carcinogenic use.

1.2 Initiate industrial and labor contacts for personal contacts for personal interviews in the counties of Hillsborough, Pinellas, Polk, and Manatee.

1.3 Initiate interviews with selected industries (and labor) within Hillsborough County.

1.4 Schedule appropriate interviews with industry in Pinellas County to be conducted during the month of November.

1.5 Convene Advisory Committee.

2.0 DURING THE PERIOD OF NOVEMBER 1, 1975 TO NOVEMBER 30, 1975

2.1 Conduct interviews with selected industries in Pinellas County.

2.2 Conduct interviews with the employees of these companies and/or the local trade unions employed predominately by these companies.

2.3 Schedule interviews with selected industries in Polk County.

2.4 Initiate compilation of data from interviews conducted in Hillsborough County.

3.0 DURING THE PERIOD OF DECEMBER 1, 1975 TO DECEMBER 31, 1975

3.1 Conduct interviews with selected industries in Polk County.

3.2 Conduct interviews with the employees of these companies and/or the local trade unions employed predominately by these companies.

3.3 Schedule interviews with selected industries in Manatee County.

3.4 Initiate compilation of data from interviews conducted in Pinellas County.

4.0 DURING THE PERIOD OF JANUARY 1, 1976 TO JANUARY 31, 1976

4.1 Conduct interviews with selected industries in Manatee County.

4.2 Conduct interviews with the employees of these companies and/or the local trade unions employed predominately by these companies.

4.3 Initiate compilation of data from interviews conducted in Polk County.

4.4 Convene Advisory Committee.

5.0 DURING THE PERIOD OF FEBRUARY 1, 1976 TO MARCH 31, 1976

5.1 Follow up interviews with industries previously surveyed if necessary and conduct interviews with any industries not previously surveyed which might be pertinent to the objectives of the project.

5.2 Analyze collected data.

5.3 Prepare and submit final report of project activities and accomplishments to F.R.M.P., Inc.

AGREEMENT OBJECTIVES

1.0 PROGRAM OBJECTIVE

To examine the extent and characteristics of industrial environmental factors which contribute both directly and indirectly to the production of cancer among employees of a selected sample of Florida industries.

2.0 OPERATIONAL OBJECTIVES

2.1 To delineate those industries and occupational groups within a four county area which are representative of the total industrial activities of the State of Florida with regard to carcinogenic exposure and health risk(s) to workers.

2.2 To identify the types of carcinogenic substances used by a selected sample of these industries and evaluate the degree of exposure hazards to employees.

2.3 To evaluate employee, industrial and occupational health activities with reference to their effectiveness among these three groups as delineated by the survey, as noted in 2.2 above.

3.0 FINAL OBJECTIVE

To ascertain the numbers of individuals exposed to carcinogenic health risks as determined by the survey and to determine the degree of success obtained in risk prevention by currently applied health precautions, and to recommend, where necessary, modifications and expansions of health precautions.

INTERIM FINAL REPORT AGREEMENT #325

****

CARCINOGENIC POTENTIALITIES OF FLORIDA INDUSTRIES

Table of Contents

Conclusions

From the results of the questionnaire it is apparent that the questionnaire technique was insufficient for eliciting the senstive data desired. Even though the cover letters clearly stated that negative as well as positive information on carcinogenic use was desired it is the opinion of the researchers that the lack of returns may stem from two causes: 1) those companies not utilizing carcinogenic substances did not see any purpose to the survey and consequently did not respond; and 2) those companies which do employ carcinogens were reluctant to respond positively for a number of reasons (potential legal culpability, immediate instigation of federal and/or state investigation, bad publicity, uncertainty regarding the confidentiality of the returns and possibily others).

Interestingly, those companies which offered the most complete returns and which have been the most cooperative in the discussion of potential carcinogenic problems are those which are considered 'giants' in their industry. This may result from 'enlightened' management and/or the financial ability to effect all known measures of prevention. Under such conditions these companies would have the least to fear from such a survey.

Of the 12 different types of industries employing carcinogenic materials, at least 6 types (aluminum milling, concrete/cement/gypsum, boat building/repairing, photographic, newspaper and plastic laminates) are highly represented among companies

employing fewer than 100 individuals. This fact alone would tend to underscore the need for intensive investigation of smaller companies which are certainly more numerous and also are perhaps less able to intitute and enforce necessary preventive measures.

As negative as this information may appear, several positive points emerge. The use of the questionnaire did reveal several types of companies which were previously not considered as users of carcinogenic materials. Knowledge of these both broadened the scope of the project and alerted the researchers to occupational groups who may be at risk. The questionnaire also demonstrated quite clearly that for obtaining information of this nature the personal interview is far superior and more effective in eliciting specifics of numbers and quality of risk management. The contrast between the apparent reluctance or inability to commit such data to paper and the more than tolerant attitude displayed during interviews is of no small importance in the achievement of the objectives of this project.

Although the researchers have maintained an unbiased attitude towards both labor and industry, quite naturally labor has taken a greater interest in this survey. It is believed that extremely good relations have obtained between the research staff and labor officials and that while still continuing an impartial posture, additional and considerably more specific data may be collected with the aid and encouragement of labor.

Recommendations

The paucity of returns from mailed questionnaires obviate specific recommendations concerning the effectiveness of carcinogenic exposure procedures, needed improvements of these procedures and/or numbers of individuals so affected. Furthermore generalizations or conclusions based upon these returns and any statistical manipulation of the results would be unwarrented and subject to considerable error.

The following recommendations are suggusted for conducting and attaining the reseach objectives in as meaningful a capacity as possible.

(1) Inasmuch as a reasonably accurate appraisal of: (a) the types of carcinogens utilized by Florida industries, (b) the number of employees exposed to these substances, (c) the present methods, programs and/or procedures for avoiding possible exposure to carcinogens, and (d) the relative effectiveness of health and safety programs and related procedures employed by industry, was not obtained by questionnaire methods, it is suggested that the research design be modified in both content and form to elicit these objectives.

(2) The methodology should be of a personal interview form. These interviews should be conducted with industrial management personnel, safety officers or personnel and employees irrespective of company size.

(3) To test the potential of the modified methodology it is suggested that such interviews be conducted in a four county area which offers both an accurate and manageable sample of the range of industrial activity within the state.

(4) The counties proposed are Hillsborough, Pinellas, Polk and Manatee which collectively account for approximately 19 percent of the state's population and also contain representatives of all industrial production within the state.

(5) That both questionnaires and interviews be given to members of unions (based upon the responses of the industrial interviews and the degree of real and/or potential exposure hazards these workers face on specific job sites) to evaluate their perceptions, attitudes and behaviors as relate to successful carcinogenic exposure prevention.

Background

The rapid proliferations of man's technology in the realms of chemistry, nuclear physics and mineral extraction has resulted in a greater intensity of exposure to both an impressive roster of synthetic and 'natural' compounds possessing carcinogenic activity and abilities. The NCI has estimated that over 80% of presenting cases of cancer in the United States are the result of environmental carcinogens introduced by man's industrial efforts. A prevalence of this magnitude strongly suggests that it is imperative that health and occupational safety personnel be particularly alert and conscious of those industries and occupational groups whose activity may present carcinogenic hazards not only to themselves, but the general public as well.

Program Objectives

The primary objective of this research project was the delineation of those industries and occupational groups in the State of Florida whose activities involves the use and/or exposure to carcinogenic substances. Additional objectives were the estimation of the number of employees so exposed, the degree of exposure (for each of the various substances), consideration of present preventive and precautionary health measures and evaluation of the effectiveness of these measures.

It was hoped that the material collected from such investigations would offer a document of substantial utility and value to diverse groups and individuals engaged in occupational health and safety. Such information would offer assistance in the design, planning and implementation of additional and/or revised policies concerning the management and control of carcinogenic substances to the benefit of all directly and indirectly concerned with the problem.

History

From the outset of this project a prime factor influencing the course and conduct of the research was the concept of the methodology to be employed. Several drafts of a questionnaire were submitted to FRMP prior to the December 3, 1974, meeting of the advisory committee. Of these, the long form (four legal pages) was presented to the committee for their advice and consideration. Discussion of the questionnaire was lengthy,

with essentially two points of view expressed. Some members thought that the long questionnaire would be rejected by the recipients as too complicated and time consuming. Members of this persuasion preferred a short, two question form which simply asked the respondent to check those carcinogens used and indicate the number of employees affected. It was noted that the short form would not be able to survey adequately the other objectives (health prevention procedures, attitudes towards such regulations as presently exist, educational programs on safety for employees and evaluation of these measures) specified in the contract. Other members of the committee were of the opinion that if the questionnaire expected to gain information regarding these objectives that a 'one shot' approach would be superior and thus, they supported the long form on the grounds that it might be more favorably received than a continued barrage of shorter questionnaires. After much discussion with the project advisory committee and F.R.M.P. staff, the long questionnaire was adopted as the survey instrument. These deliberations necessarily delayed and restricted the data collection phase of the project well beyond the time allotted.

A change in the assignment of principal investigator during the month of January, 1975, created a brief period of directional uncertainty which added to the complications of methodological considerations.

Methodology

To ascertain if similar assessments of carcinogenic material use were currently being conducted, or had recently been accomplished in other states, letters of inquiry were sent to the state public health offices of all states, excluding Hawaii and Alaska. To date 45 responses to the question have been received, all of which report negatively with regard to information of this type. Concurrently with the above, written, telephone and personal communication was conducted with the offices of National Institutes of Occupational Health and Safety, Occupational Safety and Health Administration, Florida State Health personnel and local health officials. In discussion with these people the opinion was offered that a state wide survey of any depth (on site visits, extended personal interview, expert consulting services, etc.) would not be possible given the amount of funding available. Consequently it was determined that the majority of the information would of necesscity be obtained by mailed questionnaires, with some supplementary data gathered by site visits made in the immediate vicinity of research headquarters. Although the use of a questionnaire as the survey technique was not deemed incapable of producing results, it was doubtful that a questionnaire would be the most effective means of gathering data relevant to this problem.

From the Directory of Florida Industries published by the Florida State Chamber of Commerce, 823 industries employing 100 or more persons were selected for sampling by questionnaire. To avoid

the introduction of sampling bias, no assumptions were made regarding which (if any) of the industries so selected were users of carcinogenic substances. The possible bias created by sampling according to employee numbers (perhaps a reflection of certain types of industrial activity) was considered to be negligible. A further consideration in selecting the sample on the basis of 100 employees or greater was the belief that larger companies would: 1) have both more numerous (and perhaps better educated personnel) and therefore both the chance of a return and the quality of the answers might be enhanced and 2) larger companies might have better systems of prevention and management of carcinogenic risks and would be more willing to respond with regard to this section of the questionnaire.

To encourage response each questionnaire mailed was accompanied by two cover letters, one from the researcher and the other from Mr. John C. Glenn, State Industrial Safety Administrator, explaining the purpose and necessity of the project. The questionnaire, cover letters and pretest results are included in the second progress report.

In an effort to provide complementary material from labor, it was thought desirable to solicit advice and assistance from various trade unions whose members might be exposed to carcinogenic substances. Accordingly a questionnaire similar to that sent to the industries was mailed to 48 locals comprising the Hillsborough County Central Labor Council.

Interviews within Hillsborough, Pinellas and Polk counties have been conducted with selected companies (many as part of the pretest

procedure) and labor officials. The purpose of the interviews was to augment what data was collected by the questionnaires and to test the receptiveness of both groups to personal interview.

Results of Industry Survey

Of the 823 questionnaires mailed 96 or 11.7% were returned. This low return rate was not unexpected given the broad implications of the questions that were posed. Of the 96 returns 70 companies (72.9%) indicated that they either did not use any carcinogens or were unaware of such use in either their processing and/or product(s). Nineteen industries (19.8%) stated that they did use one or more of the known or highly suspected carcinogens listed on the questionnaire. Seven companies (7.3%) declined to answer the questionnaire and responded by letter stating that their legal departments must approve all of the questions and answers or they stated that they were complying with OSHA and NIOSH regulations in addition to participating in other studies and could not devote their time and effort to this endeavor. All of the seven were either from major chemical or phosphate companies.

An additional complication regarding the validity of responses was encountered in cases of companies reporting the use of carcinogenic materials while other companies manufacturing the identical product (and according to NIOSH of necessity using identical processing) indicated no such use of these materials. On the strength of NIOSH's opinion regarding this problem 7 companies (7.3%) were added to the list of those companies

responding positively to carcinogenic use. Obviously this procedure is statistically questionable, but for survey purposes is not unreasonable. It should be noted, however, that because these 7 companies did not indicate use of these materials estimated numbers of employees exposed could not be generated. Consequently, the 7 are not included in the data summary as positive responses, but rather as negative.

The companies responding that carcinogens were utilized in their activities were of twelve different types as illustrated in Table 1.

Table 1

INDUSTRIES UTILIZING CARCINOGENS

Type of Company	Number of Companies
Aluminum Milling	1
Newspaper	1
Aerospace	3
Electrical Components	4
Citrus Processing	2
Cement/Gypsum	1
Boat Building/Repair	2
Photographic	1
Pipe (PVC) Manufacture	1
Plastic Laminates	1
Paper/Printing	1
Lasers	1

With regard to the specific types of carcinogens employed, 17 different ones, with varying degrees of exposure to 267 employees were indicated by the 19 companies' returns as displayed in Table 2.

Table 2

TYPES OF CARCINOGENS USED BY INDUSTRIES

Substance	Number of Companies	Number of Employees
alpha-Naphthylamine	1	4
beta-Naphthylamine	1	4
4.4-Methylene-bis (2-Chloroaniline)	2	8
Vinyl chloride	2	53
Radium	1	15
Uranium ores	1	15
Other radioactives	2	15
Soots	1	3
Asbestos	3	7
Chromates	8	48
Arsenicals	2	12
Nickel compounds	2	14
Auramine	1	3
Magenta	2	9
High boiling petroleum oils	1	3
Cutting oils	7	50
Creosote	2	4
		267

Although these 19 companies constituted some 20% of all companies replying to the questionnaire, it is believed that this percentage of the sample is not greater than would be reflected for industry within the entire state were all companies to be surveyed. Consultation with the Health and Safety Office at USF and personnel at NIOSH regarding the particular industries selected revealed that over half could be eliminated from consideration on the basis of both product and processing procedures. Inasmuch as the 823 companies were representative of the general spectrum of industrial activity within the state it can be said that the range and extensiveness of carcinogenic use is quite widespread.

A mean number of employees exposed, calculated on the basis of the returns indicating carcinogenic use, results in 14 employees per company. If all companies (of whatever employee size) which produce and/or process the identical materials as larger companies (those with 100 or more employees) are credited with only 1 employee who is exposed to carcinogens, then the state-wide risk factor is several thousands.

Results of personal interviews with companies operating in the Hillsborough, Pinellas and Polk county region suggest that such an estimate would not be altogether inaccurate. Of 13 companies interviewed 6 (2 chemical, 1 acid-lead battery, 1 PVC pipe, 1 metal milling and 1 paint) had employees totaling 57 who were exposed to carcinogenic hazards.

Concerning the remainder of the questionnaire which inquired about safety, availability of medical care, educational programs and corporate attitudes towards these, approximately half of the 96 returns (43.7%) completed the full questionnaire. Of those companies engaged in carcinogenic use only 15 completed the final section, either in part or entirely. The returns of these companies while indicating interest and cooperation were not of sufficient numbers to permit statistical evaluation. Responses from the 19 companies to each question are compiled in the Appendix 1.

Results from Labor Survey

Of the questionnaire submitted to 48 locals 14 responded. Most, as expected, stated that the survey was not applicable to their trade, but six locals (2 Steelworkers, 1 IBEW, 1 Chemical, 1 Meat Cutter and Butchers, and 1 Asbestos Workers) were definitely aware of the use of carcinogenic substances in their trades.

The Asbestos Workers local, comprised of some 200 individuals, is headquartered in Tampa, but includes members in a nine county area. Of all of the locals surveyed, the Asbestos Workers demonstrated the most concern and enthusiasm for the project. The researchers have been asked to four job sites and have discussed preventive and health care measures applicable to the work of insulation with the members. Further, a separate questionnaire was administered to 43 of the members during one of their local meetings. The results from this questionnaire revealed that although the members did consider themselves to be a 'risk' population they were not particularly observent of precautionary health procedures. Tabulation of the Asbestos Workers questionnaire is found in Appendix 2.

The primary safety equipment for this type of work is the partial face mask, covering only the nostrils and the mouth. Many complained that the mask filters were so thick (approximately ten layers of filter material) that they found it difficult to breathe. As a

result many refused the use of the mask on these grounds. Others expressed the opinion that the mask was worth using only when asbestos fibers and/or dust was visible. Under these conditions they were more inclined to employ their masks. This unfortunately is the reverse of what should be done, for the invisible fibers and dust are far more likely to pass through the body's natural filtering system and thus accumulate in the lungs. One last finding and one which emphasizes some of the potential cultural difficulties of adequate health and safety, was the discovery that some of the individuals did not wear their masks because they chew tobacco and consequently must stop work and remove their masks to expectorate. Of the 35 individuals responding to the questionnaire only 5 admitted to wearing their masks at all times when either working with or in close proximity to the material. This would suggest that of all trade unions surveyed the asbestos workers are far more exposed to carcinogenic risk than the others (certainly by occupation if not in numbers).

The Amalgamated Meat Cutter and Butcher Workers reported that at least 35 of their members are exposed to PVC in the course of packaging meats and other products. The PVC is in the form of the clear plastic wrapping, and in that form is not thought to be hazardous. However, the sealing of these meat packages is accomplished by melting the plastic along the bottom of the meat tray. Heating of plastic containing PVC is thought to release some of the monomer which then enters the immediate atmosphere and breathed by any one nearby. Although some members reported that masks have been issued, this procedure, insofar as NIOSH

data is considered, is totally ineffective. Further masks are not particularly comfortable and are infrequently worn by the meat cutters. It is doubtful that there is sufficient exposure to present a definite health problem in small concerns wherein only one or two people are involved in sealing. However, in larger meat departments where the sealing is continuous and several sealing machines are in use simultaneously, it is quite possible that the PVC monomer is sufficiently high to cause concern.

The United Steelworkers locals reported that at least 8-12 of their members were exposed to PVC, again in relation to a process of heating. Although they expressed concern about the PVC, they were far more alarmed about the various types of solvents and cutting oils with which they were required to work. One of the members has provided the researchers with a list of the solvents and cutting oils which was forwarded to NIOSH laboratories for identification. Should any of these prove carcinogenic then the number of affected workers could be well over 50, for the two reporting locals alone. Both returns indicated that the employees did not feel that the companies were either interested or particularly active in the area of health and safety. Incidental to the carcinogenic aspect of the survey, the researchers found that many chemical compounds of known injurious capabilities were being used by these men without adequate safeguards.

The Chemical Workers local did not respond directly to the questionnaire but did telephone to express their interest. They estimate that

approximately 30 of their members are working with PVC and other carcinogenic materials but believe that their union has investigated most of the apparent hazards. At a state-wide meeting of the AFL-CIO which the researchers attended an official of the Chemical Workers specifically requested that the researchers conduct a survey similar to that given to the Asbestos Workers. Such a questionnaire is presently being designed.

The International Brotherhood of Electrical Workers submitted a return stating that they were not aware of any carcinogenic hazard other than PVC but were concerned about some of the solvents and cutting oils used in cleaning electrical contacts. The PVC problem arises from the cutting (and melting) of PVC conduit used as electrical insulation. From this one local (employed chiefly by General Telephone) exposure to PVC could affect as many as 40 individuals. Responses from other IBEW locals engaged in similar activities would certainly elevate this figure.

From the results of the labor survey and conversations with various labor officials it is apparent that most of the members of locals are unaware of the possibility of carcinogenic exposure. This was made abundently clear by the return of the Ironworkers which stated that they did not engage in either use of carcinogens or viewed themselves subject to its presence. Interviews with labor officials of some of the locals revealed that most unions did not conduct any form of safety and health education for their members beyond mechanical concerns (e.g. keeping hands away from dangerous equipment, wearing hard hats, etc.), and further did not seem to be aware of the potential reason (depending upon the

type of trade union) for such education of both apprentices and masters.

Most of what carcinogenic information is available to the locals derives from either trade journals and/or union bulletins. However, officials of the locals do not, for the most part, insure that this information is passed to the members. Consequently many of the members when hearing of a particular hazard (e.g. PVC at the plant in Louisville) do not necessarily associate that substance with their own work. Such problems as these will no doubt continue to increase inasmuch as many of the carcinogenic materials may appear in a variety of forms. As a derivative of this problem is the hazard of 'cross contamination' or exposure. It appears that most unions are unaware of the potential risks engendered by the occupations of others. Although the Asbestos Workers are cognizant of the carcinogenicity of asbestos, this information is not generally known by electricians, plumbers, masons and others who may be working alongside (or even worse, beneath) the asbestos workers. If such groups are unaware of these hazards they will not know to take precautions and may be affected by such substances entirely innocent of using them themselves. Several instances of potential 'cross contamination' were noted by the researchers during site visits. For example, at two sites it was observed that some plumbers and electricians were working directly below open scaffolding on which cutting and shaping of asbestos insulation blankets (a process which creates considerable asbestos dust) was in progress.

APPENDIX 1

RESULTS OF QUESTIONNAIRE MAILED TO INDUSTRIES

TYPES OF COMPANIES WHICH RESPONDED NEGATIVELY TO CARCINOGENIC USE

TYPE OF COMPANY	NO. OF COMPANIES
Aerospace	2
Apparel	2
Boat windshields and tops	1
*Can manufacture	1
Can sales	1
Ceramic tile	1
*Chemical manufacture	1
Citrus	8
Communication equipment	1
Computers	2
Concrete	2
Dairy	1
Dry kilns and conveyor systems	1
Electronics	6
Engineering contractor	1
Food services, Restaurant	4
Heavy machinery	1
Jobbing machine work	1
Nuclear reactors	1
Paper, pulp, boxes, bags	9
Personal accessories	1
*Phosphate rock	2
Phosphate sales	1
Portable toilets	1
Power generating	1
Sand	1
Seafood	3
Steel products	2
Sugar	1
Windows, doors	1
Wood products, furniture	2
*Newspapers (One company only has the home office here, no manufacturing or other services)	3
Total	70

*Possibly using carcinogens, see text for explanation

RESULTS OF QUESTIONNAIRE

MAILED TO INDUSTRIES

N.B. All open ended responses refer only to those companies which stated that they used one or more carcinogens. Numbers without parentheses also refer to those companies with positive responses. Numbers in parentheses are the totals from all returned questionnaires.

2. Below is a list of substances which are known or believed to be carcinogenic (capable of producing cancer). Please check the substance which you use by listing the product and processing material(s). Also please give the number of employees who may be in direct physical contact (by breathing it, touching it, etc.). If the substance is used in more than one product or processing material, please list the substance, product and processing material on the back of this sheet.

	SUBSTANCE	PRODUCT	NUMBER OF EMPLOYEES	PROCESSING MATERIAL(S)	NUMBER OF EMPLOYEES
1-*	beta-Propiolactone	Transformers		Impregnation/Solvents	4
	4.4-Methylene-bis	Transformers		Impregnation/Solvents	4
2-	alpha-Naphthylamine	D-76		Developer	4
	beta-Naphthylamine	UFG		Developer	4
	4.4-Methylene-bis	Kodak Cleaner		Film Cleaner	4
3-	Asbestos	Stove shield	--		
	Cutting oils			Machine shop use	4
	Cresote			Protective coating	2
4-	Chromates			Metal finish	4-8
	Nickel compounds			Metal finish	4-8
	Auramine			Chem lab (8 grams)	3
	Magenta			Chem lab (5 grams)	3
	Other radioactive: 2½ & 4% Thorium in Thorium-Magnesium, Titanium Tritide foil, Americium 241.	(Not given)	--	(Not given)	-
5-	Chromates	KCr_2O_7		Waste water analyses	5
	Cutting oils	Pipe		Routine piping	10
6-	Soots			Boiler cleanup	3
	Chromates	Paints/inks		Painting/inking	12
	High boiling petroleum			#6 Fuel oil	3
	Cutting oils			Pipe threading	3

	SUBSTANCE	PRODUCT	NUMBER OF EMPLOYEES	PROCESSING MATERIAL(S)
7-	Asbestos	Insulation	3	
	Cutting oils	Cutting oils	2	
	Vinyl chloride			(Not given)
	Radium-226			(Not given)
	Uranium ores			(Not given)
	Other radioactive:			(Not given)
	Californium-252			
8-	Titium (H_3)			
	Cobalt-60			
	Chromates			(Not given)
	Arsenicals			(Not given)
	Nickel compounds			(Not given)
	Magenta			(Not given)
9-	Chromates	Fan casing		Surface treatment
	Cutting oils	Fan		Machined parts
10-	Arsenicals	Lead arsenate		Sprayed on grapefruit to hasten maturity
11-	Asbestos	Accessory product	4	
12-	Chromates			Metal conditioner
13-	Chromates	Batteries		Chromic acid (CrO_3)
14-	Chromates	Potassium dichromate		Eastman Kodak color print bleach
15-	Creosote	Boat windshields/tops	2	
16-	Cutting oils	Lasers	10	
17-	Cutting oils			Cutting aluminum
18-	4.4-Methylene-bis	MOCA (DuPont name)		(Not given)
19-	Vinyl chloride	Polyvinyl	50	

*Since many of the companies used more than one substance, the companies are and the substances used by each company follows.

CARCINOGENIC SUBSTANCES WHICH, ACCORDING TO THE SURVEY, ARE NOT IN USE AT THE PRESENT TIME:

2-Acetylaminofluorene
4-Aminodiphenyl
3.3-Dichlorobenzidine
4-Dimethylaminoazobenzene
4-Nitrobiphenyl
N-Nitrosodimethylamine
bis(Chloromethyl)ether
Chloromethyl Methyl ether
Ethyleneimine
Benzidine
Thorotrast
Coal Tars
Pitches
Tobacco
Tobacco smoke
Mustard gas
Chlornaphazine
Isopropyl oils
Shale oil
Asphalts

3. If any substances used in the processing and products are known to be cancer producing, cite those actions that your company has taken to prevent the potential health danger to employees.

1. 3 Consultation with plant physician(s).
2. 2 Consultation with plant nurse(s).
3. 4 Consultation with county public health officers.
4. 6 Consultation with NIOSH and/or OSHA
5. 0 Await some appearance of a problem to be discussed in trade or professional journals.
6. 2 Await state or federal action, e.g., the establishment of specific regulations regarding the use and management of these materials.
7. 7 Other. Please describe:

Open ended responses to number 7 (Other):

1. All employees are instructed to observe label regulations on all products.
2. Corporate safety director and corporate hygenist have surveyed process and evaluated process procedures.
3. Updating of existing ventilating and air exchange systems
4. MOCA purged from use in one section of company
5. Discussions with our corporation hygenist

6. We have no knowledge that the above materials are capable of producing cancer.
7. No action taken due to very limited exposure risk. We would welcome guidance in this area.
8. Monitoring of plant atmosphere. We have a respirator available if any high level areas are ever found.
9. Comply with OSHA standards and Air Force environmental health standards
10. Discontinue use where possible. NRC (AEC) and Florida statutory requirements for radioactive materials are follow d. Prescribe protective clothing, gear and procedural directives on others.
11. Routine precautions

4. Does your company employ a full time physician(s)?
   Yes 1 (1) No 14 (40)

5. Does your company employ a part time physician(s)?
   Yes 3 (11) No 12 (32)

6. Does your company employ a full time occupational nurse(s)?
   Yes 4 (8) No 11 (33)

7. Does your company employ a part time occupational nur e(s)?
   Yes 0 (0) No 15 (40)

8. What additional health arrangements does your company provide for employees?

1. None
2. Doctor and hospital nearby
3. Pre-employee physical and emergency ambulance service
4. Environmental health monitoring and services provided by NASA - physical examinations
5. Arrangements with local doctor to accept patients on an emergency basis, also require pre-employment physical
6. NASA furnishes all contractors with emergency medical and hospital services
7. Use industrial clinic in town to handle industrial injuries
8. Receive occupational medical support from both the Air Force and NASA medical facilities. Also carry both medical insurance and workman's comp.
9. Employees are provided with major medical insurance
10. A local doctor is used for plant assistance. Corporate H.Q. is using a clinic for possible over-all control
11. Consulting physician as needed
12. We have an arrangement with a local clinic and use of hospital facilities

13. Security guards are trained in first aid - breast self examination - coronary risk screening - tetnus shots - flu immunizations - claucoma screening - eye tests - hearing tests - drug abuse program
14. Coveralls and aprons

9. To date, has there been any discussion between your company and its employees, or representatives, e.g., union(s), regarding safety precautions and standards of exposure within the plant to known and/or suspected carcinogenic agents? If so, please describe the content and direction of such discussion(s).

1. Seven (7) companies responded "No". Of these seven two companies responded further:
   (1) No, except between management, safety director, select employees on safety committee and our insurance company
   (2) No, safety and health situations, in general, are discussed at monthly safety committee meetings.
2. N/A
3. Discussion by management as to what company is doing about VCM
4. Two letters released to advise supervision and engineering of the carcinogenic agent problem and method of control
5. Training for those working in MOCA - warning placards positioned to keep others away
6. Yes, a set guideline of procedures in handling asbestos have been incorporated and rules conduct to enforce adherence by hourly employees to the guidelines.
7. General rules enforced regarding the wearing of masks, goggles and helmets only.
8. Management/union safety committees are kept informed by safety office as necessary

10. Has your company requested any state and/or federal information about carcinogenic materials and protective procedures?
Yes 3 (5) No 13 (35)

11. Has your company received any state and/or federal information about carcinogenic materials and protective procedures?
Yes 4 (8) No 12 (31)

12. Does your company presently believe that some or all of your regulations now in effect are:

A. 0 (0) insufficient
B. 9 (12) adequate
C. 3 (3) stringent
D. 0 (0) too stringent
E. 1 (2) too stringent, and applied before a problem was proven to exist.
F. 0 (0) not worth the cost of their enforcement
G. 0 (0) other response to regulation(s); please list:

13. Which protective measures and regulations in the opinion of your company are not effective?

1. Three (3) companies responded "N/A"
2. Material to e ucate employees
3. The use by OSHA of the chemical substance rather than a trade name and number
4. The low level (.5 ppm) which could cause an area to become regulated

14. How could these be improved?

1. Three (3) companies responded "N/A
2. Location and acquiring such materials
3. By increasing to 25 ppm on a time weighted average
4. By requiring a carcinogenic warning label on all products as well as the trade name and number designation in addition to the chemical substance.

15. If you can or wish to be specific, which regulations do you believe your would wish to be dropped?

1. Three (3) companies responded "N/A"
2. None, just include better labelling

16. Could you outline why these regulations should be dropped?

1. Four (4) companies responded "N/A"

17. Does your company believe that the costs for protective measures should be borne by:

A. 1 (9) the company
B. 2 (2) the government (state and/or federal)
C. 1 (5) the public through product price increase
D. 7 (13) some combination only if the cost is extremely high, otherwise by the company

18. Has your company initiated any safety precautions not explicitly required by OSHA regulations?
    Yes 7 (14) No 9 (16)
    Please describe them if response is yes:

    1. Too many to list
    2. In plant first aid station, full time nurse
    3. A controlled hazardous substance list is maintained to prevent such commodities from undetected use
    4. We are regularly inspected by a state official and have had a recent OSHA inspection
    5. All employees in accessories dept. wear approved paper masks and protective coveralls at all times when in dept.
    6. One letter to supervision and engineering to advise what the safety precautions would require to obtain safety approval of MOCA potting compound
    7. Probably - we have been in operation longer than OSHA -- materials which might be similar to suspected carcinogens but not listed would get the same precautionary treatment.

19. What segment of your employee population do these precautions affect?

    1. Entire work crew
    2. All
    3. 80%
    4. Lab and non-destruct testing personnel - potting and mold shop mostly. All generally
    5. Four employees at maximum
    6. None present
    7. N/A

20. Has your company experienced any difficulty in having the employees observe protective measures which you have instated?
    Yes 4 (13) No 11 (18)
    If yes, please describe the difficulty:

    1. There are always those who refuse to comply
    2. Will not observe pressure limits on air outlets. Careless in wearing respirators
    3. Getting employees to wear the personal protective equipment, and then getting them to wear this equipment correctly.
    4. Protective equipment such as safety glasses take getting used to -- complaints at beginning
    5. We have some areas of high decible noise. Ear plugs were purchased and distributed to the individuals concerned. As of date, I know of no individual using them.
    6. All hazardous work areas require safety signature approval
    7. N/A

21. Has your company sponsored any educational programs to alert your employees to hazards of improper protection or exposure to dangerous materials or substances (i.e., talks by representatives of your public health department, American Cancer Society, etc.)?

Yes 8 (13) No 6 (19)

Please describe the types of programs which you offered and its effect upon the employees.

1. Internal program through safety committee with information supplied from safety office. Keep employees informed -- also on consumer products safety commission bans.
2. Only through our safety director and management stressing the importance relative to the employee health and well being
3. Foremen train employees in use of materials. They enforce safety regulations
4. Discussions with supervisors, safety posters, discussion at safety committee meetings. Most employees conform, no effect on "renegades".
5. All employees were given a class on VCM and its potential effects. It was well received.
6. Safety poster program. Foremen level. Safety pamphlets. First aid training program.
7. Launch operations employees receive certification and recertification in the hazardous disciplines encountered in their work tasks in addition to safety workmanship meeting.
8. N/A

22. Considering industries in general, does your company believe that the risk of cancer to employees is sufficient to justify governmental establishment and enforcement of regulations?

Yes 3 (10) No 9 (18)

Comments:

1. Not a problem here
2. N/A

APPENDIX 2

RESULTS OF QUESTIONNAIRE GIVEN TO ASBESTOS WORKERS UNION

1. Age 18-64 (32.7*)

2. Sex All Males

3. Single 8; Married 25; Divorced 2; Widowed ____; Remarried ____.

4. Education: Less than 9th grade 3; 9th-10th grade 2; High school 19.
1-2 years college 7; 3-4 years college 1; College degree 1.

5. Where were you born? ____ City and ____ State.

6. Did you grow up in: a small town 16; medium size city 10; large city 9.

7. Did your father come from: a small town 23; medium size city 8; large city 3.

8. Did your mother come from: a small town 23; medium size city 8; large city 3.

9. Were your parents union members? Father 15; Mother 2; Both ____.

10. How many years have you worked with insulation materials? .5-35 Years. ((11.6*)

11. Have you had any respiratory problems to date? Yes 4 No 30.

Did you see a doctor? Yes 4 No ____.

What was the respiratory problem diagnosed as? Asbestosis (2); Chemical Pneumonia

Have you had any reoccurrence? Yes 2 No 2.

Were you satisfied with the correctness of the diagnosis? Yes 3 No ____.

12. Do you regularly see a doctor for a general checkup? Yes 12 No 21.

13. When was the last time you were admitted to an emergency room or hospital? ____

What was the reason? ____

14. Do you smoke cigarettes? Yes 12 No 23. If yes, how many packs per day? 1.4*

How long have you smoked cigarettes? 3-25 (13.2*)

Have you ever smoked? Yes 9 No 8. If yes, when did you stop? ____ years ago

When you quit, how many packs per day were you smoking? 1.7*

When did you first begin smoking cigarettes? 12-23 (17*) years of age.

15. Have you ever worked with an individual who suffered lung damage or other health problems from working with insulation materials? Yes 20 No 13.

*Mean value for responses to question.

16. On the job, do you think that other occupational groups (electricians, plumbers, etc.) are exposed to much of the insulation material which you are working with? Yes 22 No 10 . Do you think that these other occupational groups are aware of the possible hazards of insulating materials? Yes 8 No 23 .

17. Do you think that any of the insulation materials with which you work are dangerous to your health? Yes 31 No 1 . If yes, which one(s) Asbestos (13) Adhesives (13), Fiberglass (20) .

Which of the following was most influential in helping to form your opinion of the danger of these substances? Fellow workers 21 Radio-T.V. 6 Newspapers 6 ; Union health literature 16 ; Trade journals 12 ; Your personal doctor 5 ; Other .

18. Have you ever suggested to workers of a different trade that they move away from your work area to prevent their being exposed to some of the materials that you were using? Yes 27 No 6 .

19. Do you ignore certain safety procedures? Yes 20 No 12 . If yes, do you ignore them because they make you physically uncomfortable? Yes 16 No 4 .

20. When was the last time you saw a doctor? .

21. Do you have any relatives who also work as insulators or roofers? Yes 15 No 17 .

22. To what extent do current safety procedures hinder you in your work?
Very much ; considerably 7 ; very little 16 ; not at all 8 .

23. What do you consider to be the most common violation of safety procedures that are specific to your trade? Respirator not worn(16), Inadequate ventilation (5) Safety equipment not furnished (1) .

24. Which of the insulation materials with which you have worked do you find to be the most difficult or uncomfortable to work with even though it presents no health hazard? Fiberglass blanket (13), Mineral wool blanket (7), Epitherm (5) .

25. Do you work fulltime? Yes 27 No 1 . This averages about how many hours per week 40* .

26. How effective do you think your union at the national level is in making efforts for health and safety procedures for the members? Very effective 4 ; effective 10 ; somewhat effective 13 ; not at all effective 4 .

27. Do you have specific suggestions about health and safety procedures which you would like to see practiced? Yes 13 No 12 . If yes, what? Enforced mask use (5) Better ventilation and dust collection (2), Enforce present protective laws (3), Dampen materials to prevent dust (2), Safety meetings (2), Alert other crafts to health dangers (1).

28. Would you be willing to answer a more detailed questionnaire? Yes 28 No 2 .

)

MEMBERS OF THE ADVISORY COMMITTEE

Charles Cowl
FRMP Advisory Board (Labor)
1701 North Franklin
Tampa, Florida 33602

Robert Hebblethwaite
Division of Health
1217 Pearl Street
Jacksonville, Florida 32201

William Hill, M.D.
Polk County Health Officer
Winter Haven, Florida 33880

Richard S. Hodes
State Representative
238 E. Davis Blvd.
Tampa, Florida 33606

Granville W. Larimore, M.D.
Director
Florida Regional Medical Program
One Davis Boulevard, Suite 307
Tampa, Florida 33606

John Shebel — Salser - Heiman (TECO)
Associated Industries of Florida
119 N. Monroe
P.O. Box 784
Tallahassee, Florida 32302

PROJECT STAFF:

Travis J. Northcutt, Jr., Dean
College of Social and Behavioral Sciences
University of South Florida
Principle Investigator

Ailon Shiloh, Ph.D.
Professor of Anthropology and Director of Graduate Studies
University of South Florida
Principle Investigator

Robert I. Gilbert, Jr.
Research Associate
University of South Florida

Grace Oliver
University of South Florida
Mrs. Oliver will be taking minutes of our meeting

AGENDA

I. Opening Remarks

II. Discussion of Project Objectives

III. Discussion of Contract

IV. Closing Remarks

ADVISORY COMMITTEE MEETING MINUTES

The first advisory committee meeting for Carcinogenic Potentialities of Florida Industries, FRMP Grant No. 325, was held December 3rd, 1974, at 2:00 p.m. Present were Mr. Charles Cowl, Mr. Robert Hebblethwaite, Dr. William Hill, Messers William Salzer and Andrew Heman, acting on behalf of Mr. John Shebel, Dr. Granville Larimore, Mr. Spero Moutsatsos, Dean Travis Northcutt, Dr. Ailon Shiloh, and Mr. Robert Gilbert. Representative Hodes was unable to attend. Mr. Cowl asked that Mr. Morris Blake, Executive Director of the Building Trades Council be informed of the proceedings of the project and considered as an alternative for labor representation. Dr. Shiloh called the meeting to order. Members of the advisory committee were introduced by Mr. Gilbert.

Dr. Shiloh began with a statement of the purposes of the grant and discussed the obligations in meeting and the three operational objectives stipulated by the contract. The operational objectives to be met are:

1. To identify industries and occupational groups in Florida where there is exposure to carcinogenic materials as to location, number of exposed workers, type of exposure, and related demographic and epidemiological characteristics;

2. To identify the types and degree of hazard of carcinogenic material to which the workers are exposed;

3. To review the industrial hygiene and occupational health activities directed at the prevention of occurrence of cancer among these workers, and their effectiveness.

Dr. Shiloh pointed out that although the project staff entailed four people, the actual research was, in fact, being carried on by one person. Dr. Shiloh then called for a report of the progress being made in each of the delineated time periods specified in the contract.

Mr. Gilbert reported on the first quarter, and passed the progress report around. It was suggested that the progress report be made available to all

members of the advisory committee. Mr. Moutsatsos stated that FRMP would take responsibility for duplication of the report. Dr. Shiloh then raised the question of procedure by which the industries employing carcinogenic material might be delineated. The methodology for delineating these industries is crucial to the direction of the project. Dr. Shiloh suggested that a questionnaire could be sent to all industries inquiring of their use of known carcinogens. This would be a major screening and depending on the returns, further follow up might be by additional questionnaires, telephone, letters, or on site visits.

Messrs Heman and Salzer commented that many industries might not be aware of which chemicals or agents used were carcinogenic and suggested that a list of carcinogenic materials be attached to the questionnaire. An intermediate possibility was suggested, in that the industries could be sampled and/or selected on the basis of numbers of employees and/or products. This method, particularly if done by employee size alone, might allow for oversights with industries that do handle carcinogenic agents, but who have a very small employee population, thus permitting an industry employing say 50 people with minimum exposure to be included whereas 10 similar industries with 10 employees each might be excluded. The difficulties inherent in this approach were discussed by Dr. Hill and Mr. Moutsatsos. Mr. Gilbert suggested that another alternative might be selection of industry solely by product (and if possible knowledge of intermediate processing) derived from the Directory of Florida's Industries. These industries would be sent questionnaires, contacted by phone, and, if possible, followed up by site visits. Debate of the three approaches continued for some length.

Mr. Gilbert passed a copy of the questionnaire, which had been made up earlier, and discussion ensued regarding the applicability and/or potentially threatening questions which it contained. Regarding the questionnaire, Messrs

Salzer, Heman, and Cowl were of the opinion that if the questionnaire could be made extremely brief and direct this would improve chances of completed returns. Mr. Moutsatsos, Dr. Northcutt, and Dr. Hill felt that if this pattern were followed, it was possible that too much information might be lost and were less inclined to the 'watered down' approach.

Messrs Moutsatsos and Cowl suggested the possibility of a regional, in depth, study as opposed to a state wide study. While recognizing the merits of such an approach, such a stradegy would entail redefinition of the contract as it now exists. Mr. Cowl urged that the research definitely include consultation with workers, safety committees, and management, but that these be conducted separately, both in terms of questionnaires and face to face contact. Dr. Northcutt concurred in the separation of such investigation for not only would it permit imput from both sides, but would enhance the objectivity of the survey.

The potential legal difficulties entailed by on site visits were considered. Both Dr. Larimore and Mr. Moutsatsos felt it might be possible for us to make these visits in conjunction with OSHA officials at the time that they were making their own visits. Messrs Cowl and Salzer both commented that we had no legal right to conduct such on site visitations and, therefore, the study could be considerably hindered in its objectives. Discussion ensued regarding what procedures could be employed to insure both the confidentiality of any information gained from industry and methods which might obtain the confidence of industry. This topic concluded with no concrete solutions, but recognition that the actual stradegy of industrial selection would to a large extent determine approaches to on site visits.

Dr. Shiloh discussed the fact that Mr. Gilbert had conducted a survey of all state public health offices inquiring if they had done or were currently contemplating doing a similar state survey and the response from all replying

to date (over thirty) was negative. Dr. Shiloh emphasized that this attempt is a pioneer effort and argued for maintaining a state wide approach. Messrs Salzer and Moutsatsos suggested in response to this that perhaps we target in on five industries and follow these in depth. Mr. Gilbert responded that we should consider the question of whether our concern should be directed towards the number of employees involved (statewide) as opposed to concentration on industries (perhaps with small work forces) which do use highly potent carcinogenic agents. Mr. Gilbert introduced this not with the idea of exclusion of one or the other, but the two different concepts (numbers of workers versus carcinogenic potentency) were being subsumed under the catch phrase "high risk". The general tenor seemed to be that both should be investigated, once again raising the question of the basis and methods of industrial selection.

Sources of carcinogenic information were considered with the following suggestions: Mr. Gilbert might be sent to discuss this with officials at NIOSH and ICR; that lists of carcinogenic materials might be obtained from regional offices of OSHA and NIOSH; and additional sources might be obtained from state divisions of health and labor. Mr. Gilbert reported that attempts had already been made with NIOSH, OSHA, and state divisions of health and labor. Although expressing interest and cooperation, these agencies were unable to provide more specific information that the list of 14 known agents published in the Federal Register. Dr. Hebblethwaite concurred in recognizing the difficulty of finding industries using these agents. He stated that the state did not possess this information.

Dr. Hill emphasized that whatever approach we followed in obtaining and publishing carcinogenic information, it should be done extremely carefully to avoid creation of any public alarm similar to that occasioned by the "cranberry scare". He also stated that admirable as a state survey may be,

we should design the project around what could be most effectively accomplished in terms of "useful" information whether that dictated the regional or other limiting approach.

Dr. Shiloh had to leave the meeting for classroom obligations, but in concluding emphasized the desirability of obtaining a statewide view of the problem. He was of the opinion that the operation could be conducted in two stages: some form of generalized questionnaire which should elicit the industries using carcinogenic agents (stage 1), and then following up those industries replying in the affirmative to carcinogenic use by letter, phone, or in person (stage 2).

Recognizing both the difficulty in either a mass mailing or moving from product to industrial selection, Dr. Larimore and Mr. Moutsatsos offered to fund some students to assist with either approach. Mr. Gilbert responded with appreciation for this offer, but suggested that a person with professional chemical expertise would be more desirable if industrial selection is based upon the product and its intermediate processing. Dr. Hill then left the meeting due to clinic commitments.

Dr. Northcutt concluded the meeting with the suggestion that the project staff meet with Dr. Larimore and Mr. Moutsatsos to determine what direction data collection and industrial selection take. Members of the advisory committee were advised that following the meeting with FRMP, they would be notified of the conclusions and given ample time to respond with suggestions or criticisms to those conclusions. It was emphasized by the project staff that the purpose of the advisory committee is not for "rubber stamping", but rather to offer any advice, suggestions and/or criticisms which will enhance the success of the project.

Voicing the appreciation of the project staff for the time and effort which they devoted, Dr. Northcutt adjourned the meeting at 5:10 p.m.

CARCINOGENIC POTENTIALITIES OF FLORIDA INDUSTRIES

ADVISORY COMMITTEE MEETING

DECEMBER 3, 1974

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

AND

THE UNIVERSITY OF SOUTH FLORIDA

TO: FINANC... ...UNTING

DATE: SEP 6 1974

SUPPLEMENT TO AFFILIATION AGREEMENT

BETWEEN

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

AND

University of South Florida
Carcinogenic Potentialities of Florida Industries

This supplement to Affiliation Agreement is made this 1st day of August, 1974, by and between FLORIDA REGIONAL MEDICAL PROGRAM, INC., (hereinafter F.R.M.P.) and University of South Florida, (hereinafter the Recipient).

WHEREAS, F.R.M.P. and the Recipient are parties to an Affiliation Agreement dated the 1st day of July, 1974, and F.R.M.P. and the Recipient wish to enter into a supplemental agreement contemplated by the aforesaid Affiliation Agreement in order to carry out F.R.M.P.'s operational program Activity No. 325, entitled Carcinogenic Potentialities of Florida Industries.

NOW, THEREFORE it is agreed that all provisions of the Affiliation Agreement between F.R.M.P. and the Recipient dated the 1st day of July, 1974, shall apply to this supplement agreement as if fully set forth herein. The schedules and exhibits attached hereto are a part of this Supplemental Agreement. The period of this Supplemental Agreement will be from August 1, 1974 to July 31, 1975.

The following schedules form a part of this Agreement:

Schedule No.	Page No.	Schedule Title
No. 1	3	Objectives
No. 2	4,4A	Activities
No. 3	5	Evaluation
No. 4	6	Performance Reports
No. 5	7,7A1,A2,B,D,F,G	Budget
No. 6	8	Reimbursement
No. 7	9	Expense Reports
No. 8	None	

Date: August 1, [illegible]

Schedule No.	Page No.	Schedule Title
No. 9	None	
No. 10	"	
No. 11	"	
No. 12	"	
No. 13	"	
No. 14	"	
No. 15	"	
No. 16	"	
No. 17	"	
No. 18	"	
No. 19	"	
No. 20	10	Termination and Extension

In witness whereof, the parties hereto have hereunto excuted this Agreement for the purpose herein expressed the day and year above written.

Signed, sealed and delivered in the presence of:

Barbara Copson

Karen Swaney

As to F.R.M.P.

FLORIDA REGIONAL MEDICAL PROGRA

By: J. Phillip Hampton
President

_____

_____

As to Recipient

University of South Florid
Recipient

By: Cecil Mackey

Name:
Title:

APPROVED AS TO FORM AND LEGALITY
ATTORNEY - U. S.

FRMP-OP-40

Schedule No. 1
Agreement No. 3
Date: August 1,

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Agreement Objectives

The objectives of this Agreement are those described below. Any change in objectives must have the prior written approval of F.R.M.P., Inc.

1.0 PROGRAM OBJECTIVE:

To determine the extent and characteristics of the environmental factors which play a significant and perhaps a major role, both directly and indirectly, in the production of cancer in human beings with special reference to the carcinogenic potentialities of Florida's industries.

2.0 OPERATIONAL OBJECTIVES:

2.1 To identify industries and occupational groups in Florida where there is exposure to carcinogenic materials as to location, number of exposed workers, type of exposure, and related demographic and epidemiological characteristics.

2.2 To identify the types and degree of hazard of carcinogenic material to which the workers are exposed.

2.3 To review the industrial hygiene and occupational health activities directed at the prevention of occurrence of cancer among these workers, and their effectiveness.

3.0 FINAL OBJECTIVE:

To ascertain and delineate the existence and range of cancer control programs among high risk industries and occupational groups in Florida.

Schedule No. 2
Agreement No.
Date: August 1

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Schedule of Activities

The activities to be undertaken by the Recipient are those activities described on this page. Any change in the scope of these activities must have the prior written approval of F.R.M.P., Inc.

1.0 During the period of August 1, 1974, to October 31, 1974:

1.1 Employ project staff, two co-directors, one research assistant, and one temporary part-time typist.

1.2 Survey existing literature to provide pertinent baseline data.

1.3 Establish contacts with state agencies; e.g., Division of Health, Departments of Commerce and Agriculture, medical and hospital associations, manufacturing associations, labor unions, mining industries, food producers, in Florida to obtain pertinent reports and data.

1.4 Make arrangements with the Florida State Division of Health to provide data on incidence and morbidity from cancer of the lung, skin, and any other forms of cancer of possible industrial or occupational origin in 1950, 1960, and 1970.

1.5 Seek sources of reference materials on the prevention of occurrence of cancer in industrial and occupational groups with special reference to Florida industries.

2.0 During the period of November 1, 1974, to December 31, 1974:

2.1 Delineate those selected industries and occupational groups to be studied.

2.2 Select and convene an advisory committee* with representation from state and local governments, industry and labor, medical and hospital associations, state and local health officers, industrial hygienists and occupational health experts and consumer representatives.

2.3 Develop and pretest the survey instrument design to be used for data gathering and review with F.R.M.P., Inc., and the advisory committee.

3.0 During the period of January 1, 1975, to April 30, 1975:

3.1 Delineate the work schedule to be followed by the staff for the duration of the project.

3.2 Complete the epidemiological analysis from a sample of death certificates from cancer of the lung, skin, and any other forms of cancer that might be associated with industrial exposure.

3.3 Initiate data collection through in-depth studies and on-site visits of known industries and occupational groups exposed to carcinogenic materials using the pretested survey instrument.

* To include Granville W. Larimore, M.D., Director, Florida Regional Medical Program, as an Ad Hoc member.

Schedule No. 2
Agreement No. 32
Date: August 1,

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Schedule of Activities

The activities to be undertaken by the Recipient are those activities described on this page. Any change in the scope of these activities must have the prior written approval of F.R.M.P., Inc.

4.0 During the period of May 1, 1975, to July 31, 1975:

4.1 Complete collection, analysis, and interpretation of all pertinent data assayed.

4.2 Initiate and complete writing and editing final draft of the final report.

4.3 Submit the final draft of the final report to the advisory committee.

4.4 Prepare and submit the final report to F.R.M.P., Inc.

Schedule No. 3
Agreement No. 325
Date: August 1, 19

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Evaluation

F.R.M.P. staff, consultants supported by F.R.M.P., the Florida Regional Advisory Group and others which F.R.M.P. may authorize from time to time, may evaluate progress of the activities to be undertaken as set forth in schedule No. 2 page 4. Such evaluation will be based on the criteria set forth in schedule No. 3 page No. 5. The Recipient and personnel employed for the purpose of this contract will cooperate in facilitating such evaluation of performance.

The F.R.M.P., Inc. staff, advisors and consultants will evaluate the project at regular intervals that are agreeable to both parties. This will be done by appropriate evaluation techniques such as site visits, technical reviews and special studies of performance and outcomes.

The performance report schedule will be one of the processes used to assess the performance of the project and its activities.

Schedule No. 4
Agreement No. 325
Date: August 1, 197

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Performance Reports Schedule

Recipient will file periodic reports of performance for the periods and by the dates set forth below. Such reports will contain detailed information as shown below for each report.

1.0 Initial progress report* will be filed by November 15, 1974, for the period of August 1, 1974, to October 31, 1974:

1.1 Curriculum vitae of project staff.

1.2 Summary of initial literature search.

1.3 Report of contacts made with state agencies; e.g., Division of Health, Departments of Commerce and Agriculture, medical and hospital associations, manufacturing associations, labor unions, mining industries, food producers, in Florida to obtain pertinent reports and data.

2.0 Second progress report* will be filed by January 15, 1975, for the period of November 1, 1974, to December 31, 1974:

2.1 Delineate those selected industries and occupational groups to be studied.

2.2 Report on the pretest of the survey instrument design.

2.3 Report on the advisory committee meeting including minutes of the advisory committee meeting.

3.0 Third progress report* will be filed by May 15, 1975, for the period of January 1, 1975, to April 30, 1975.

3.1 Delineate the work schedule to be followed by the staff for the duration of the project.

3.2 Report detailing the data collection phase.

4.0 Fourth and final report**will be filed by August 31, 1975, for the period of May 1, 1975, to July 31, 1975:

4.1 Minutes of the advisory committee meeting.

4.2 Copy of final report with overall findings and recommendations.

* 5 copies
** 10 copies

Agreement No. [illegible]
Date: August 1, 1974

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Budget Schedule

F.R.M.P. will provide funds as reimbursement for eligible expenses. Eligible expenses are only those shown in the budget which forms a part of this schedule as pages 7A through 7G. Funds may not be transferred between line items without the prior written approval of F.R.M.P. Forms to request the transfer of funds between line items will be provided by F.R.M.P. With respect to the total budget and amounts for each line item, F.R.M.P. may at its sole discretion increase or decrease the amounts shown. Should the total budget be increased or decreased, other schedules forming a part of this agreement with respect to objectives (Schedule No. 1) and activities (Schedule No. 2) will be subject to modification.

Schedule No. 5
Agreement No. 325
Date: August 1, 1974

Budget Period
From: August 1, 1974
To: July 31, 1975

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Budget Schedule

Quarterly Operating Budget

BUDGET CATEGORY	1st Quarter Ending 3/31	2nd Quarter Ending 6/30	3rd Quarter Ending 9/30 1974	4th Quarter Ending 12/31 1974	TOTALS
Salary and Wages			$1,750	$ 2,625	
Employee Benefits			24	36	
Movable Equipment					
Consultant Costs			100		
Supplies			150	100	
Domestic Travel			100	150	
Rental of Space					
Rental Other Than Space					
Publication Costs					
Contractual Service					
Communication Costs			143	144	
Computer & Data Processing					
Trainee Stipends					
Other Trainee Costs					
Other Costs					
TOTAL DIRECT COSTS			$2,267	$3,055	
INDIRECT COST			$1,011	$1,517	
TOTAL DIRECT & INDIRECT			$3,278	$4,572	

Schedule No. 5
Agreement No. 325
Date: August 1, 1974

Budget Period
From: August 1, 1974
To: July 31, 1975

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Budget Schedule

Quarterly Operating Budget

BUDGET CATEGORY	1975 1st Quarter Ending 3/31	1975 2nd Quarter Ending 6/30	1975 3rd Quarter Ending 9/30	4th Quarter Ending 12/31	TOTALS
Salary and Wages	$ 2,625	$ 2,625	$ 875		$10,500
Employee Benefits	36	36	12		144
Movable Equipment					
Consultant Costs	100				200
Supplies	100	100	50		500
Domestic Travel	150	100			500
Rental of Space					
Rental Other Than Space					
Publication Costs			200		200
Contractual Service					
Communication Costs	117	123	51		578
Computer & Data Processing					
Trainee Stipends					
Other Trainee Costs					
Other Costs					
TOTAL DIRECT COSTS	$ 3,128	$ 2,984	$ 1,188		$12,622
INDIRECT COST	$ 1,517	$ 1,517	$ 505		$ 6,067
TOTAL DIRECT & INDIRECT	$ 4,645	$ 4,501	$ 1,693		$18,689

Schedule No. 3
Agreement No. 325
Date: August 1, 19

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Itemized 12 Month Budget for Personnel

The total for salary shown on this page must be exactly equal to the total salary and wages total shown on page 7A. The total for benefits shown on this page must be exactly equal to the total employee benefits shown on page 7A.

Name & Discipline	Job Title or Function	Institutional Affil. (Inc. Subunits, i.e. Depts., Schools, etc.	Time or Effort % hours	Amt. Authorized (Omit Cents)		
				Salary	Benefits	Total
Travis J. Northcutt, Jr., Ph.D.	Co-Director	University of South Florida	10%	- 0 -	- 0 -	- 0 -
Ailon Shiloh, Ph. D.	Co-Director	University of South Florida	10%	- 0 -	- 0 -	- 0 -
Robert Gilbert, M.A.	Research Assistant	University of South Florida	100%	$ 9,600	$ 132	$ 9,732
TBA	Temporary part-time typist	University of South Florida	100%	900	12	912
TOTAL				$10,500	$ 144	$10,644

7

FRMP-OP-47

Schedule No.
Agreement No. [illegible]
Date: August 1, 19

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Consultants

Describe by name the consultants to be employed, the compensation to be received by each and the purpose of their employment. Employment of consultants must be documented in accordance with the provisions of the Regional Medical Program: Guidelines. The total shown below must equal the amount budgeted for consultants on page 7A.

	Consultant's Name	Purpose of Employment	Amount
1	TBA	For travel, meals and lodging related to advisory committee meetings only*	$200
2	TBA		
3	TBA		
4	TBA		
5	TBA		
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

Total $ 200

* To include travel only: no fee to be paid

Schedule No. 5
Agreement No. 325
Date: August 1, 197

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Domestic Travel

List persons authorized to travel on Agreement funds, purpose for which travel is authorized and the location to which travel is authorized.

Name of Traveler	Purpose of Travel	Location Authorized
1) Travis J. Northcutt Jr. Co-Director	Development and implementation of project	State of Florid
2) Ailon Shiloh Co-Director	Development and implementation of project	State of Florid
3) Robert Gilbert Research Assistant	Data Collection	State of Florid

TOTAL $500

FRMP-OP-51

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Budget Detail: Other Cost

Items of cost that are not described on pages 7A through 7F of Schedule No. 5 must be described on this page. If funds are shown in the "Other Cost" line item, describe each item to be purchased.

SUPPLIES	$ 500
Postage and supplies	
PUBLICATION COSTS	200
Publication of Final Report	
COMMUNICATION COSTS	578
Telephone	
TOTAL	$1,278

Schedule No. 6
Agreement No. 325
Date: August 1, 1974

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Reimbursement

F.R.M.P. will reimburse eligible expenses as they are presented on F.R.M.P. expense voucher forms OPV-1 and OPV-2. These forms are attached as Exhibits A and B of this Agreement. The Recipient will file expense vouchers monthly. Reimbursement will be made as vouchers are received provided expenses reported conform to the budget set forth in Schedule No. 7, are eligible costs under the Guidelines: Regional Medical Programs, are consistent with the activities described in Schedule No. 2 and provided that at the time expense vouchers are received all progress reports then due have been received and approved by F.R.M.P.

An advance of funds not to exceed twenty percent of the direct cost budget of this contract will be made upon execution of this agreement. Therafter this advance shall be charged against reimbursement so as to reduce the advance to a zero balance at the end of the term.

Schedule No. 7
Agreement No. 325
Date: August 1, 1974

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Schedule of Expense Reports

(1) During the period of this contract, the Recipient will file expense reports on a quarterly basis as follows:

Report for the Period	Due
January 1 through March 31	April 20
April 1 through June 30	July 20
July 1 through September 30	October 20
October 1 through December 31	January 20

This quarterly report will compare expenses to budget on the form to be provided by F.R.M.P. and will include funds obligated and actual cash disbursements.

(2) If this Agreement terminates on December 31 of any year, the Recipient will file with F.R.M.P. a list of Agreement cash disbursements and funds obligated by line item on the form to be provided by F.R.M.P. This report will be filed within ten (10) days of the end of the agreement period. If it is necessary to file this report, then it will not be necessary to file the report described in paragraph No. 1 due January 20.

(3) Ninety(90) days after the termination of this contract, the Recipient will file a final expense report for the Agreement on the form to be provided by F.R.M.P. If at this date there are accrued but unpaid expenses, a list of such expenses including names and amounts, will be furnished F.R.M.P.

Schedule No. 20
Agreement No. 325
Date: August 1, 1974

FLORIDA REGIONAL MEDICAL PROGRAM, INC.

Termination and Extension Provision

This Agreement may be cancelled upon thirty (30) days written notice by F.R.M.P. or the Recipient. Such cancellation will not, however, affect the duty of the Recipient to account to F.R.M.P. pursuant to the Affiliation Agreement of July 1, 1974, nor affect the duty of the Recipient to file reports of expenditures and progress as may be applicable at the date of cancellation. In addition, continuation is subject to the availability of grant funds provided by the United States Department of Health, Education, and Welfare.

All reports and notices to be given under this Agreement shall be mailed to the parties at the following indicated addresses:

FLORIDA REGIONAL MEDICAL PROGRAM, INC.
1 Davis Boulevard, Suite 307
Tampa, Florida 33606

University of South Florida
Recipient
4202 E. Fowler Avenue
Tampa, Florida 33620

UNIVERSITY OF SOUTH FLORIDA
TAMPA • ST. PETERSBURG

COLLEGE OF SOCIAL AND BEHAVIORAL SCIENCES
DEPARTMENT OF ANTHROPOLOGY
TAMPA, FLORIDA 33620

813: 974-2138
813: 974-2140

November 26, 1974

Dr. William Hill
Polk County Health Officer
Winter Haven, Florida 33880

Dear Dr. Hill:

This is to confirm the date for the advisory committee meeting on December 3rd at 2 p.m. at the Social Sciences Building, Room 285.

As I indicated in my telphone conversation we are still in the planning stages of collecting the information for meeting the following objectives:

1. to identify any industries or occupational groups which expose employees to carcinogenic materials as to location, number of workers so exposed, type of exposure, and related demographic and epidemiological characteristics;

2. to identify the types of carcinogens to which employees are exposed and develop a control program(s) utilizing current technological and scientific information;

3. to review present industrial hygiene and occupational health activities designed for the prevention of carcinogenic exposure and evaluate their effectiveness.

The meeting on December 3rd is called to invite your suggestions for methods by which we could most effectively fulfill these objectives and for additional points of concern regarding the topic which you may share.

I have enclosed a visitor's parking pass for your convenience. If you should forget it, please stop by the information booth upon entering campus and they will provide you with another.

Thank you for your time and effort on our behalf and I look forward to a productive and stimulating meeting.

Sincerely,

Robert I. Gilbert
Research Assistant

UNIVERSITY OF SOUTH FLORIDA
TAMPA • ST. PETERSBURG

COLLEGE OF SOCIAL AND BEHAVIORAL SCIENCES
DEPARTMENT OF ANTHROPOLOGY
TAMPA, FLORIDA 33620

813: 974-2138
813: 974-2140

November 25, 1974

Dear Advisory Board Member:

This is to confirm the date for the advisory committee meeting on December 3, 1974 at 2:00 p.m. at the Social Sciences Building, Room 285.

As I indicated in our telephone conversation, we are still in the planning stages of collecting the information for meeting the following objectives:

1. to identify any industries or occupational groups in which there is exposure of employees to carcinogenic materials, as to location, number of workers, type of exposure, and related demographic and epidemiological characteristics;

2. to identify the types of carcinogens to which employees are exposed and develop control program(s) utilizing current technological and scientific information;

3. to review the present industrial hygiene and occupational health activities designed for the prevention of carcinogenic exposure and evaluate their effectiveness.

The meeting on December 3rd is called to invite your suggestions for methods by which we could most effectively fulfill these objectives, and for additional points of concern regarding the topic which you may share.

I have enclosed a visitor's parking pass for your convenience. If you should forget to bring it, please stop at the information booth at the entrance to campus and they will provide you with another.

Thank you for your time and effort on our behalf, and I look forward to a productive and stimulating meeting.

Sincerely,

Robert I. Gilbert
Research Assistant